1-1-2021

Developing Computer Aided Algorithms For Automatic Assessment Of Psoriasis Disease Severity Using Digital Images

Intisar Rizwan I Haque

[How does access to this work benefit you? Let us know!](#)

Follow this and additional works at: https://commons.und.edu/theses

Recommended Citation

Rizwan I Haque, Intisar, "Developing Computer Aided Algorithms For Automatic Assessment Of Psoriasis Disease Severity Using Digital Images" (2021). *Theses and Dissertations*. 4183.
https://commons.und.edu/theses/4183

DEVELOPING COMPUTER AIDED ALGORITHMS FOR AUTOMATIC ASSESSMENT
OF PSORIASIS DISEASE SEVERITY USING DIGITAL IMAGES

by

Intisar Rizwan I Haque
Bachelor of Engineering, National University of Sciences and Technology, 2009
Master of Science, National University of Sciences and Technology, 2012

A Dissertation

Submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Grand Forks, North Dakota

December
2021

Name: Intisar Rizwan I Haque

Degree: Doctor of Philosophy

       This document, submitted in partial fulfillment of the requirements for the degree from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

Jeremiah Neubert

Ryan Adams

Colin Combs

Bryon Grove

Ivan Lima Jr.

       This document is being submitted by the appointed advisory committee as having met all the requirements of the School of Graduate Studies at the University of North Dakota and is hereby approved.

Chris Nelson
Dean of the School of Graduate Studies

12/8/2021

Date

PERMISSION

Title	Developing Computer Aided Algorithms for Automatic Assessment of Psoriasis Disease Severity Using Digital Images
Department	Biomedical Engineering
Degree	Doctor of Philosophy

In presenting this dissertation in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my dissertation work or, in his absence, by the Chairperson of the department or the dean of the School of Graduate Studies. It is understood that any copying or publication or other use of this dissertation or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my dissertation.

<div align="right">

Intisar Rizwan I Haque
December 2021

</div>

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I wish to express my sincere appreciation to the members of my advisory Committee for their guidance and support during my time in the doctorate program at the University of North Dakota.

ABSTRACT

The purpose of the research work was to develop image processing algorithms for automatic and objective assessment of psoriasis to aid in more effective management and treatment of patient symptoms. For this purpose, multiple classical machine learning-based unsupervised approaches are proposed which achieve a maximum average accuracy of ~94% on a dataset of 45 images of various resolutions. This accuracy improves performance by 14% in comparison to existing approaches reported in the literature on the same dataset of images. An alternative deep learning neural network-based approach was proposed that achieves an average accuracy of ~87% for the same set of images with minimal pre and post processing.

Additionally, a mechanism to classify images into red, red + scaling and scaling classes is proposed based on the color profile of the lesion that provides an objective way of measuring the severity of the redness of the lesion and the scaling. Apart from the model development, a survey of existing mobile apps was carried out to determine the need for this application along with the technological gaps. Based on this survey and the interviews with various stakeholders associated with the psoriasis disease treatment, a pathway to commercialization is proposed by developing a complete business model.

INTRODUCTION

Psoriasis is an incurable, complex, chronic, and inflammatory skin disease that causes raised, red, and scaly patches to appear on the skin. To improve the quality of life of patients living with psoriasis, a computer aided assessment tool that provides automatic objective assessment of the disease severity is proposed in this dissertation. The computer aided algorithms are based on classical machine learning approaches with both supervised and un-supervised learning as well as deep learning approaches. The proposed approach using k-means clustering and another approach based on superpixels (unsupervised learning) was able to achieve an accuracy of ~92% and ~94% respectively. While Support Vector Machine (SVM) based approach achieved an accuracy of (~84%). The deep learning neural network-based approach was able to achieve an accuracy of (~87%) using the same dataset of images. The advantage of using the proposed deep learning-based approach is minimal pre and post processing required for the images and does not require any feature engineering. The additional contribution of this work includes a mechanism to sub-classify images based on lesion characteristics into red, red + scaling and scaling classes. This allows for the quantitative determination of presence of scales inside the lesion based on the color profile of the lesions. The three classes of lesions have different color characteristics therefore require different dominant color features to be extracted for machine learning model development. Also, it provides a convenient way to qualitatively describe the characteristics of the lesion based on presence and absence of the erythema and scaling inside the lesion. Additionally, a pathway to technology commercialization is provided through development of a complete business model;

1

a survey of existing mobile apps related to psoriasis disease was carried out to categorize the various apps available across popular app stores based on functionality and usage.

Psoriasis affects 2% to 3% of the world's population [1] and an estimated 7.5 million people in the United States alone [2]. Psoriasis patients often suffer from severe physical pain, discomfort, social isolation and stigmatization, and psychological distress. The costs associated with this incapacitating disease are estimated at $11.25 billion per year [3], and psoriasis patients pay an average of $2,528 out-of-pocket expense annually [4]. Because of high treatment costs, health insurance providers compel dermatologists to thoroughly document the care provided, meaning more time is spent per patient than for other diseases. Current methods to monitor and assess psoriasis are limited to visual inspection and the care provider's sense of touch, which results in inconsistent assessments among clinical professionals such as nurses or dermatologists [5]. The Centers for Disease Control and Prevention (CDC) have identified the lack of objective measurement of disease severity as one of the priority areas in addressing the public health agenda for psoriasis [6]. Keeping in view the significant impact of psoriasis on health-related quality of life (HRQoL), disease severity and assessment of patient-reported outcomes (PROs) is considered vital for defining effective treatment options [7]. In addition, physicians need to prepare a thorough clinical documentation of disease severity to prove the medical necessity for the prescribed medications to obtain approval from the insurance. The process requires a lot of time-consuming paperwork due to expensive medications costing an estimated $26,708 per year, on average, and may be as high as $67,148 annually for certain medications [8]. Such high costs

often lead to payers requiring a step therapy approach for the patients where they must first use low-cost medication and if it does not help than incrementally shift to higher cost but more effective medications. Patients, on the other hand, do not have any assessment tools available to keep track of disease progression and thus rely on visual perception. But as response to medication often takes time ranging from three to six months during which time new flare-ups of psoriasis may also occur. This makes it difficult for the patients to notice if the treatment is working which often leads to non-adherence of therapy. Additionally, as psoriasis at present does not have a cure, pharmaceutical companies keep conducting clinical trials for new effective therapies, but the absence of automated assessment tools require complex, time consuming, and costly clinical protocols to prove the efficacy of the medications for Food and Drug Administration (FDA) approval.

Thus, the current need for consistent and objective assessment of psoriasis led this research to develop a computer aided approach that detects the skin regions inside the image and removes background from the image, then features (such as color and texture information) are extracted so that the image can be segmented to determine the ratio of the lesion to normal skin; that is the percentage of the body surface covered with psoriasis, also known as body surface area (BSA) measurement. In addition, the skin redness level and severity of the presence of white scales are estimated to determine the approximate psoriasis area and severity index (PASI) score. PASI is a clinically validated psoriasis scoring method approved by the American Academy of Dermatology (AAD) to quantify, track, and differentiate various stages of the disease. Currently, BSA and PASI are estimated visually by dermatologists.

Measurements of psoriasis lesions based on clinically validated severity assessment scales remain an open area of research.

The potential benefits of this research and its subsequent commercialization include improvement in existing treatment workflow for clinical professionals that will enable automatic treatment documentation acceptable for insurance approval for medications, thus reducing time spent per patient by an estimated 50% ($\approx$10 minutes). This, in turn, is expected to reduce insurance approval time for medications from an estimated average of 3 days to a few minutes, thus benefitting psoriasis patients. The availability of a clinically validated, objective severity assessment tool will help patients keep track of disease progression to improve their understanding of treatment efficacy, hence increasing adherence to therapy. Also, the development of such a tool will ease clinical trial data reporting for FDA approval of new treatment drugs by research and development scientists at pharmaceutical companies.

CURRENT STATE OF THE ART

Mobile Health Apps

Patients suffering from chronic conditions like psoriasis require regular checkup visits to physicians. Treatment of the disease often becomes difficult when patients are not located geographically near to the treatment site. For instance, the distance between patient and physician could prohibit regular visits, as may be the case for people living in rural areas, or poor weather conditions could make travel impossible [9]. For such cases, the widespread usage of smartphone technology along with the availability of high-speed mobile broadband internet, even in remote areas, provides opportunity for development of mobile health (mHealth) solutions. Even though no standardized definition of mHealth exists, the Global Observatory for eHealth (GOe) defines mHealth as medical and public health practices supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants (PDAs), and other wireless devices. The devices primarily use core mobile services like voice and short messaging service (SMS), 3G and 4G mobile internet, global positioning system (GPS), and wireless connectivity such as Bluetooth [10].

In the year 2012, more than 97,000 mHealth applications, which primarily focused on general health and fitness, were available for download in app stores across the smartphone platforms [11]. The number of apps available since then has increased many fold. During 2015 alone, almost 100,000 mHealth apps were developed by 58,000 mHealth App publishers, with more entering the market with each passing day [12]. The lack of regulation and no requirement for clinical validation for mobile apps has slowed mHealth app integration with physicians and

healthcare providers. An existing review paper details the mHealth taxonomy through a detailed survey of mobile apps, but it discusses the broader applications of mHealth within the healthcare system from the point of view of point-of-care diagnostics, patient monitoring, wellness, behavior modifications, therapy compliance and instructional applications [13]. Keeping in mind this wide proliferation of mHealth apps in recent years, a review of existing apps and their features for managing various aspects of psoriasis disease were carried out to determine the application of these apps and potential gaps that still exists.

The management of psoriasis can be improved through mobile apps providing such features as

- logging of symptoms and triggers,
- providing educational resources about psoriasis,
- analyzing patient diets with psoriatic triggers automatically,
- providing advice for minor lifestyle changes through analysis of past behavior that can prevent frequent flare ups,
- recording medication,
- providing medication reminders and information about medications and their possible side effects,
- recording changes in disease severity over time through image analysis and medication tracking to estimate treatment efficacy and improve personalized care,
- recording family history to determine the risk of psoriasis onset in healthy individuals, and

- providing offline remote physician advice and live remote interaction between the physician and the psoriasis patient to minimize office visits.

In the following section of this chapter, first, a review of literature about the existing criteria of mobile app evaluation is discussed; second, the research methodology and results are provided; third, comments from dermatologists and patients are summarized which were obtained through interviews; fourth, a discussion on the barriers to the widespread usage of mobile apps and current technological gaps in the treatment of psoriasis are presented; and finally, concluding remarks.

Background on mHealth Applications

The vast availability of mHealth apps related to a range of patient health care applications provides a diverse set of features and greater selection for end users. But the problem is how to define a standard metric for evaluation of the quality of the mobile apps for patients to improve their mobile app usage experience. Generally, mobile apps provide a customized user interface for an optimized access to content with the intent that the apps respond quickly, utilize minimal mobile data, and consume less energy [14]. These set of parameters define the quality of experience for app users. Various attempts have been made to develop mHealth evaluation criteria which have been based on multiple factor such as: 1) user reviews for (a) ease of use, (b) reliability, (c) quality, (d) scope of information, and (e) aesthetics [15]; 2) criteria based on app life cycle (a) development, (b) implementation, (c) integration, and (d) sustained operation [16]. These criteria certainly measure various aspects of app quality, but the problem is that these rating scales are not implemented within the various app stores like Apple iTunes or

Google Play Store. The existing available rating scales use the star ratings ranging from 0 to 5 stars and are published on retailers' web pages or the respective app store. Apart from the rating scale, information about the number of user downloads of the app and app reviews are available for assessing the app quality and popularity among the users. Higher star ratings, greater number of app downloads and installation generally represent greater usage and higher popularity of the app among the users.

Methodology for Systematic Review of Mobile Apps

A systematic search of the apps was conducted with the search term 'psoriasis' on several app stores including Android, iOS, Amazon, Windows, Blackberry, Galaxy apps, Firefox, and Ubuntu, considering the PRISMA guidelines for systematic literature reviews [17]. An exhaustive list of psoriasis related mobile apps was created in a spreadsheet. The approach is shown in Figure 1.
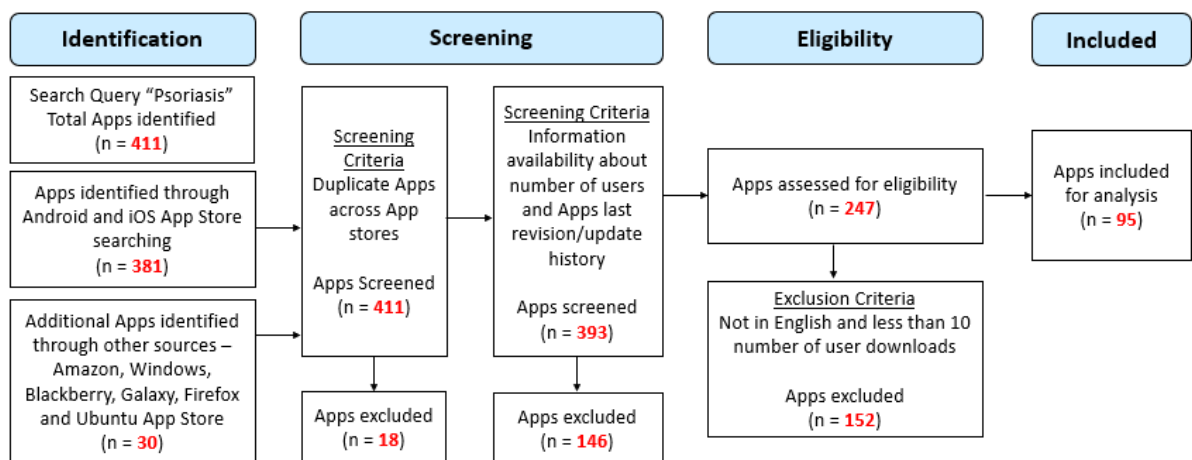
Figure 1. Systematic Review Approach of Existing Mobile Apps Related to Psoriasis Based on PRISMA Guidelines

As shown in Figure 1, a total of 411 apps were found with 249 apps on Android, 132 apps on iOS, 21 apps on Amazon, 3 apps each on Windows, Blackberry and Galaxy, and no results from the Firefox and Ubuntu app stores. The apps available across multiple app stores were considered as a single app. Only applications that related to psoriasis were considered in this work. Additionally, apps with fewer than 10 downloads, or that were not available in English were excluded from consideration.
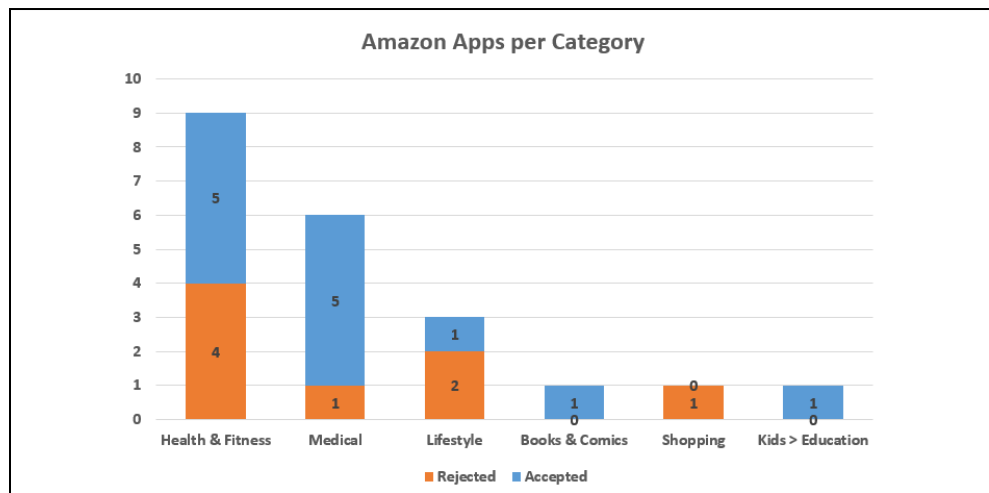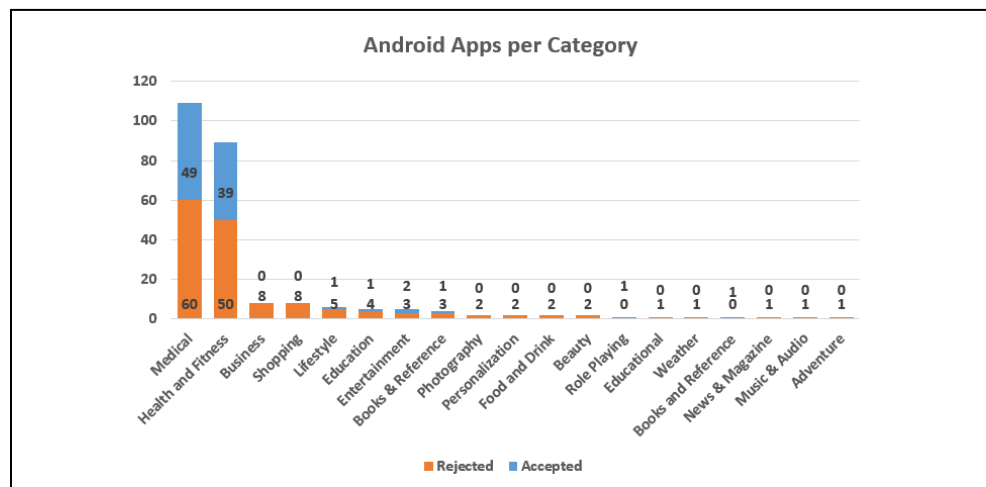
Figure 2. App Categories

For each app, the number of users as a measure of user preference for the app features, the last revision/update of the app as a measure of the level of support from the developer to reflect new medical information and the category of the app was then recorded. These categories were supplied by the respective app stores and are summarized in Figure 2. Figure 2 shows the apps per category that were accepted and rejected for the major app stores. The major categories of psoriasis apps are related to medical, and health & fitness apps, which together account for over 70% of each major app store.

The remaining apps (95) were then classified based on their core functionality into 9 distinct categories:

(i)   Contact Directory

Contact directory apps provide information about the non-profit patient advocacy organizations, resources, and support for patients to live successfully with their skin conditions.

(ii)   Clinical Reference

Clinical Reference apps are primarily intended for physicians as a guide for diagnostics. These apps are typically hard to understand for non-physicians.

(iii)   Educational Resources

Educational apps are intended for developing the understanding of psoriasis. These apps include information on causes and symptoms, the difference between normal and diseased anatomy of skin, prevention, treatment, and common complications. This information may be

presented through text, pictures, animated videos, and real videos. Sometimes advice regarding self-care and healthy lifestyle for prevention of psoriasis is also included.

(iv)   <u>Loggers</u>

Logger apps keep track of symptoms and triggers, map trends, and display progress and results through various visualization tools like graphs and charts. In addition, they may provide medication reminders or help monitor changes in a user's condition with a library of user-taken images. They may also provide advice on improving skin condition through slight changes in daily life.

(v)   <u>Social Networking</u>

Social networking apps enable people living with psoriasis to connect with other psoriasis patients, especially those living nearby, to share each other's experiences, treatment results, and how they cope with psoriasis in their daily lives. Like other popular social networking sites (example Facebook), these sites provide the ability to post daily updates as well as provide and receive comments of support.

(vi)   <u>Survey Apps</u>

Survey apps are designed to collect data on the happiness of people living with psoriasis, the impact of medication on psoriasis treatment, and the triggers for psoriasis. These apps usually require users to fill in short questionnaires with frequency varying from once daily to once monthly. The contributions are usually collected anonymously and pertain to user experiences and opinions regarding psoriasis.

(vii)   Psoriasis Assessment Calculators

Psoriasis Assessment Calculators are primarily intended for dermatologists to calculate Psoriasis Area Severity Index (PASI) scores and Body Surface Area (BSA). These two tools are universally used for assessment of psoriasis severity. In addition, the apps may provide the ability to create patient records, patient pictures databases, or records of previous PASI scores to track the patients' conditions. In addition, some apps also assess the stress and anxiety level of the patients with the Dermatology Life Quality Index (DLQI) questionnaire.

(viii)   Physician Contact

Physician Contact apps allow users to send images or questions to board certified physicians. These apps may be used by patients to contact a specific physician or by anonymous individuals to contact one of several board-certified physicians. These apps usually include some method to send images to the physician and often provide a feature to schedule appointments.

(ix)   Home Remedy

 Home Remedy apps offer users a collection of home treatments for many conditions. These can include a variety of methods such as diets, exercises, or topical skin treatments. These apps usually do not mention medication or provide information about medications.

Results

The number of apps that provide each of the above nine core functionalities, along with the approximate number of users who downloaded and installed these apps on to their mobile

devices is shown in Table 1. The data was gathered through Google Play store only because the Amazon store and iTunes store does not provide app installation information.

Table 1: Comparison of Apps Within Each Category and Number of Downloads										
Total Number of Users App Download										
Total	**33**	**21**	**18**	**9**	**5**	**3**	**3**	**2**	**1**	**95**
100,000 - 500,000	0	0	0	1	0	0	0	0	1	**2**
50,000 - 100,000	1	0	0	0	0	0	0	0	0	**1**
10,000 - 50,000	1	3	0	1	0	0	0	0	0	**5**
5,000 - 10,000	3	2	4	0	1	0	0	0	0	**10**
1,000 - 5,000	4	6	5	2	2	3	0	2	0	**24**
500 - 1,000	2	2	2	1	0	0	1	0	0	**8**
100 - 500	5	3	4	2	0	0	2	0	0	**16**
50 - 100	4	1	1	1	0	0	0	0	0	**7**
10 - 50	13	4	2	1	2	0	0	0	0	**22**
Download Range	Physician Contact	Home Remedy	Educational Resource	Logger	Clinical Reference	Psoriasis Assessment Calculator	Contact Directory	Social Networking	Survey App	**Total**
Category of Apps										

As shown in Table 1, Logger, Physician Contact, Home Remedy, and Educational Resource apps accounted for 85% of the apps. Each of the Logger, Survey, and Physician Contact apps category had more than 50,000 users.

Another important aspect for Mobile App evaluation is to assess the update history of the app. The update history indicates level of support from the app developer, resolution of software bugs found and reported by users, and compatibility with new versions of the Mobile OS (iOS/Android). In addition, for mHealth related apps, the continuous updating of medical information is highly critical since this information can change dramatically over time.
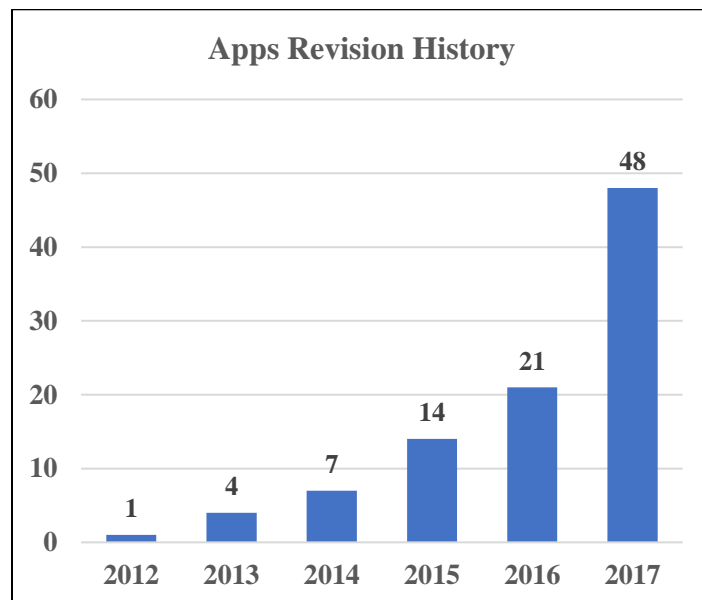
Figure 3. Apps Update History

Figure 3 indicates that less than one fourth of the 95 apps were updated during the current year and less than half the number of apps were updated in the last sixteen months.

Dermatologist and Patient Perspective

To identify the key problems and challenges associated with the treatment of psoriasis and the involvement of various stake holders within the healthcare landscape, 126 interviews were

conducted, ranging from clinical professionals, psoriasis patients, pharmaceutical company representatives, insurance companies, and pharmacies.

For the psoriasis patients, the key challenge identified was the treatment as each patient wants their skin to be clear from the lesions which most often does not happen as there exists no cure for psoriasis and the treatment only manages the symptoms. Patients find it hard to believe that a cure is not available despite the significant scientific advancements. As the patients start to realize the life-long nature of the ailment, they tend to give up on the treatment because of the fear of side effects, financial and family problems, forgetfulness, or lack of visible improvement over time, resulting in problems in the patient–physician relationship. In fact, the prime reason for the unsuccessful treatment is linked to discontinued therapy because of a lack of motivation in the patients. Patients lose confidence in the treatment and the care provider as response to medications often takes time, ranging from 3 to 6 months, during which time new flare-ups of psoriasis may also occur. And in case of new flare-ups, it takes at least 3 weeks to see any noticeable changes with the treatment. This makes it difficult for the patients to notice the treatment response and keep track of disease progression solely through visual perception due to lack of severity assessment tools. When the patient perceives that the medications are not working, it leads to frustration and often leads to non-adherence. In fact, research indicates about 60% of the psoriasis patients are shown to have poor adherence to therapy [18]–[20]. Manifestation of the disease causes social embarrassment to the patient as others treat it as a contagious disease, resulting in low social acceptance and depression.

An additional challenge for the patients is keeping track of the medications taken. The treatment requires the use of a combination of biologics and topical therapies, with varying dosage need, that is adapted for each patient as per disease location and severity. Therefore, patients often need to account for either the number of pills left after few days to check if they have been correctly following the medication routine or try other alternatives. Patients often rely on home remedies or try to maintain a manual log of daily activities identify potential triggers causing more flare-ups. This strenuous effort is fruitful only after long trial and error. As a result, most patients usually give it up as a futile exercise.

For the dermatologists, key challenges associated with the psoriasis treatment are 1) getting insurance approval for prescription medicine and 2) ensuring adherence to therapy by the patients. Insurance plays one of the most important roles in determining the type of treatment that the patient receives as it dictates what medications are covered. Clinical documentation of disease severity is required to prove the medical necessity for the prescribed medications for obtaining approval from the insurance through a process known as prior authorization. The process requires care providers to adopt step therapy approach for prescribing medications. Even if care providers know that the initial medications may not be as effective for the treatment, care providers still need to follow the requirements set by the insurance to provide documented clinical confirmation.

The criteria for failure of treatment medication are minimal improvement observed in BSA over 3 to 6-month period. To obtain approval from insurance, physicians are required to provide a history of medication usage by the patients and documented failure of each

medication used. The documentation becomes more complicated if the patient has seen multiple providers, in which case the patient needs to have prior medication records indicating what he/she has been treated with in the past. Otherwise, the similar treatments might be repeated, and the patient may have to wait an additional 3 to 6 months before shifting to new medication. At times, providers may need to call prior treating physicians to ascertain the medications prescribed. Additionally, prior authorization may be denied even after submission of clinical documents because of a lack of coverage for the medication, in which case another medication needs to be prescribed, leading to more paperwork.

Few dermatologists take pictures of the psoriasis lesions in addition to the clinical notes which is important for making a stronger case for approval from insurance. In addition, these pictures are shown to the patients on the follow-up visit to let them see the improvements in skin condition since their last visit to build confidence in the treatment.

Lastly, Body Surface Area (BSA) is used in clinical practice even though Psoriasis Assessment and Severity Index (PASI) is a more accurate measure of psoriasis severity and provides the means to check treatment efficacy. The reason is dermatologists consider the speed of assessment to be more important as compared to the accuracy of the severity assessment. As PASI is more complicated and time consuming compared to BSA, it is not used.

Barriers to Usage and Development of Mobile Apps

Based on the interviews that we conducted, we found out that most dermatologists do not take pictures because of time constraints as they need to spend 5 to 10 minutes per patient for documenting the disease condition through clinical notes, which includes the BSA estimated through visual assessment, often resulting in inter- and intra-physician variability. The lack of publicly available image resources has severely limited the development of technology that will enable automatic assessment and monitoring of disease progression over time through smartphones.

The existing mobile apps do not provide automatic correlation between symptoms and triggers and the images of the psoriasis lesions which require patients to deduce on their own what triggers are causing more flare-ups. This may be challenging for most common users lacking scientific or technical knowledge.

Additional issues include the lack of trust of the health care providers on the diagnostic accuracy of the mobile apps due to lack of clinical trial validation, lack of citations for the content within the educational related apps, and the lack of involvement of physicians or other medically trained staff in the development of apps. A lot of apps are plagiarized through repackaging of existing apps resulting in security vulnerabilities due to presence of malware or virus inside in the app.

Chapter Conclusion

Even though thousands of mHealth apps are widely available for a variety of health conditions, the number of apps related to psoriasis are relatively small. The available psoriasis apps, once filtered and classified based on their primary functionality, indicated the lack of any software for assessment of disease severity through image analysis or for assessment of treatment efficacy with medication. In addition, results also point to lack of long-term continuous support from the existing app developers, which in turn will affect the functionality of the app on new mobile devices and OS. However, the active participation of users who downloaded survey apps indicates the strong interest of psoriasis patients in mHealth apps. To ensure the creation of high-quality apps, the input of psoriasis patients must be included during the development of future apps. The dermatologists should try to take pictures to document the progression of disease condition over time. This will increase the availability of disease specific image databases thus enabling development of more automatic identification, monitoring and assessment of disease severity over time. The need of the hour is to develop a single app for psoriasis patients with well-integrated features across the various functions mentioned in this chapter.

PSORIASIS TREATMENT – HEALTHCARE LANDSCAPE AND COMMERCIALIZATION

Commercialization of the technology under development is an important milestone that guided the direction of research at various stages of current research. In this chapter, business model development for the current research is presented in detail for the possible future commercialization of the technology.

Initially 21 customer discovery interviews with psoriasis patients, dermatologists, nurse and insurance agents were completed as part of the National Science Foundation (NSF) funded I-Corps ZAP! and BOOM courses offered by the Innovation Node - Los Angeles National Science Foundation under grant number 1444080. The purpose was to perform customer discovery research to identify the value proposition and required features for a minimum viable product while the technology was under initial stages of development to make necessary changes to align with customer needs. Based on the initial interviews, an initial hypothesis was made that the nutritionists, estheticians, R&D scientists at pharmaceutical companies, patients living with psoriasis and their dermatologists and physicians were the possible customer segments of technology. The value proposition for the psoriasis patients was anticipated to include: 1) The ability to self-document in a way that provides consistent and reliable calibrated image measurements and clinically accepted Psoriasis Area and Severity Index assessment; 2) Tele-consultation access to dermatologists or estheticians for cosmetics advice or second opinions; 3) Personal disease management tools such as medication reminders, appointment scheduling systems and dermatologist contact directories, symptom, trigger, diet

tracking, and treatment efficacy analysis; 4) The ability to share anonymous treatment information and recovery response between patients with similar psoriasis conditions through secure social networking.

While the value proposition for clinical professionals was anticipated to include: 1) A reliable record of a patient's disease provided through patient self-documentation, and/or reliable in-clinic image based documentation of psoriasis; 2) Automated and consistent PASI tracking, and documentation of symptoms, trigger, diet tracking, and treatment efficacy analysis; and 3) Tele-consultation access to patients who live outside of urban centers and ability to provide consultations to patients who may be located in other regions, providing additional sources of revenue. The value proposition for nutritionists and estheticians included access to potential customer segments, without limitation of geographical access.

For the research and development scientists, the value would be clinically valid severity assessment tools to best report their clinical trial data in their submissions for FDA approval. Medical insurance providers were considered as potential influencers of this commercialization effort. The distribution of the technology to the customer segment was to be made through the mobile app stores and the dedicated online website. The customers recruited for clinical data collection were to be provided with the app for initial assessment and review with included benefits for referral marketing. Patient advocacy groups (e.g., National Psoriasis Foundation), American Academy of Dermatology (AAD) publications and

dermatology atlas websites (e.g., Dermnet) were considered for marketing and social media access to dermatologists and nurses.

The required key resources were to include: 1) image processing experts; 2) technology development grant funding; 3) application developers; 4) dermatologists; 5) public datasets; and 6) IP protection. Key activities were to include: 1) developing image processing algorithms; 2) developing cloud-based analysis algorithms; and 3) clinical data acquisition, analysis, and validation. Main partners include Spectral Molecular Imaging (SMI), Inc. Sherman Oaks, CA and the State of North Dakota (ND), Department of Commerce. The revenue model for different customer segments was to be subscription service, pay as per need or in specific cases "freemium." The cost structure was to include costs of 1) App development and overhead; 2) licensing of cloud infrastructure; 3) clinical trials; and 4) regulatory costs. The technology under development was branded as TAMpso (**T**elemedicine based **A**ssessment and **M**anagement of **pso**riasis). The initial business model canvas is shown in the Figure 4.

Soon in to the I-Corps program, we learned that our initial value propositions were in fact features of the technology under development. These features, in our opinion, were solutions to the problems for the various stake holders involved in the treatment of psoriasis. Based on our initial 5 interviews and the input received from the I-Corps program instructors we quickly revised the value proposition and reduced the number of customer segments to include only the psoriasis patients and the dermatologists. The value proposition for the psoriasis patients was to save time by reducing the appointment time from 1 month to 1 day. As we learned that patients need to wait on average a month to see a dermatologist.

(a)

Value Proposition

For Psoriasis Patients: Save time by reducing the appointment time from 1 Month to 1 Day
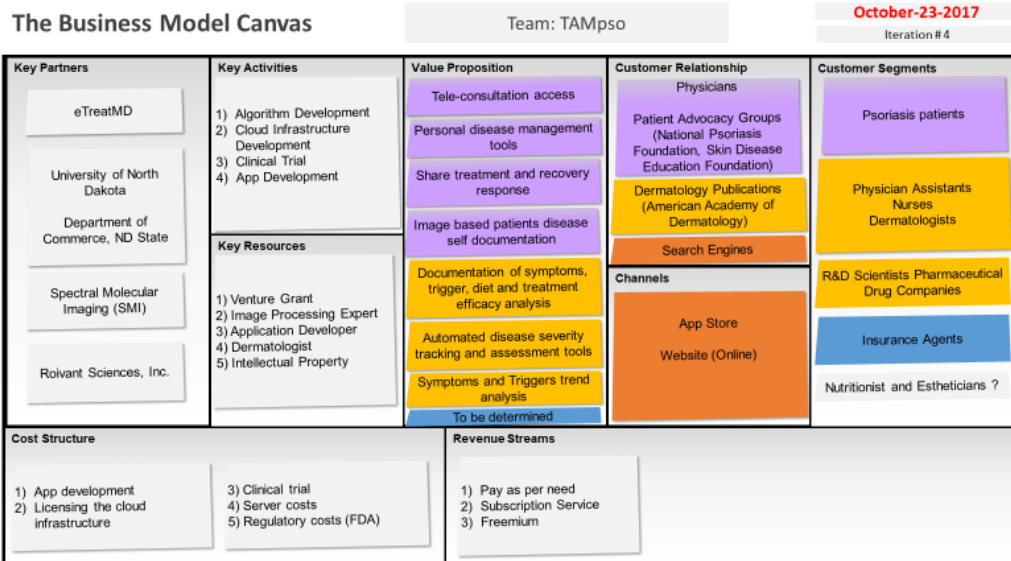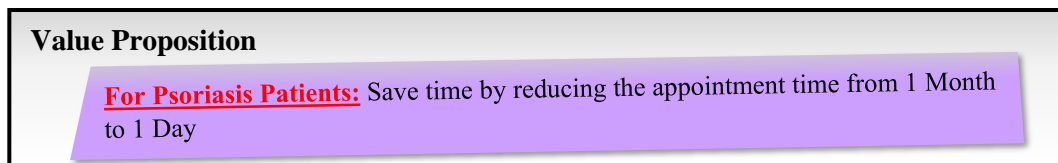
(b)

Figure 4. (a) Business model canvas at the start of I-Corps program (b) Revised Value Proposition

Subsequent 8 interviews with dermatologists, nurse practitioners, medical assistant and psoriasis patient revealed that appointment wait time varied from 1 week to 6 weeks (average 3 weeks) and physicians spent 15 to 30 minutes per patient during the in-clinic patient visit. Additionally, psoriasis is visually assessed to see the severity of the disease and the use of PASI Score instead of Body Surface Area (BSA) currently used in clinical practice will not have much effect on treatment. Pictures of the skin condition are taken for follow up

comparison of the treatment effectiveness. Based on this information value proposition was revised as shown in Figure 5.
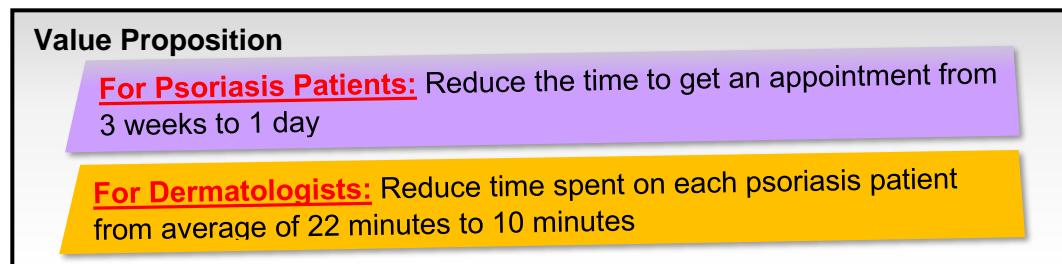
Figure 5. Value Proposition Revision 2

Further 10 interviews with dermatologists, nurse practitioners, medical assistants and psoriasis patients revealed that reducing the time to get an appointment from 3 weeks to 1 day is not important as dermatologist can accommodate patients who are in urgent need of care. Additionally, the dermatologist pointed out that the use of tele-medicine may be beneficial for the patients for the follow up visits but preferred not to use it for the first visit due to inability of the physician to touch the skin which is essential for diagnosing the skin diseases. Dermatologists suggested that same charges be applied for the tele-consultation and in-clinic visit. This led us to discard the value proposition for reducing appointment time for the psoriasis patients.

Additionally, the triggers for disease flare-ups vary from one patient to another. Information about triggers can help in adjusting the treatment if for example for some people winter season causes more flare-ups. Such knowledge can help in adjusting medicine during winter or if one is travelling to the cold areas. Patients also rely on home remedies or try to maintain a manual log of daily activities to identify potential triggers causing more flare-ups. This strenuous effort is fruitful only after long trial and error. As a result, most patients usually

give it up as a futile exercise. Insurance companies require that the patients follow a step therapy approach where they must first use low-cost medication. If the patient does not respond to these medicines over a certain time, then treatment is incrementally shifted to higher cost but more effective medications. The reason for step therapy is the costly treatment associated with the psoriasis with costs as high as $5,400 per month for prescription injection. Dermatologists pointed out that improving adherence to therapy to control its progression is extremely important as the disease can develop into more serious conditions like arthritis and even cardiac problems. Based on this information value propositions were revised as shown below.

Value Proposition

For Psoriasis Patients: ~~Reduce the time to get an appointment from 3 weeks to 1 day~~

For Psoriasis Patients: Improve understanding of psoriasis treatment and disease triggers

For Dermatologists: Reduce time spent on each psoriasis patient from average of 22 minutes to 10 minutes

For Psoriasis Patients and Dermatologists: Comprehensive tracking of relation between triggers, skin changes and treatment

Figure 6. Value Proposition Revision 3

Based on the information thus gathered and the above value propositions, led us to develop the initial customer workflow diagram shown in Figure 7. As shown in the Figure 7, the workflow starts with the patient making an appointment with the dermatologist through TAMpso mobile app. The patient enters the perceived disease symptoms and acquire images of the lesions through TAMpso mobile app.
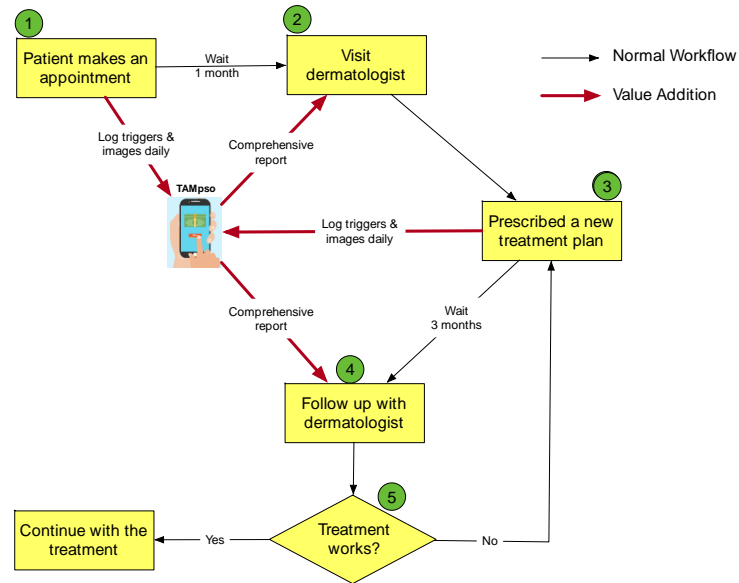
Figure 7. Customer Workflow Diagram

During the month-long waiting time before the appointment, any changes in the disease condition are logged in the mobile app. Than patient visits the dermatologist who will have access to the comprehensive disease report. Based on the report and patient examination, the dermatologist will prescribe a treatment plan for the patient. Subsequently, the patient will go through the treatment for three months while keeping the log of images. Surprisingly, three months is the usual follow-up time for the dermatology patients. During the follow-up visit, the comprehensive report will help objectively point out disease changes and its progression. Based on this information, dermatologist will either continue the treatment or decide to change to a new treatment plan.

The information gathered through these interviews was used to draw the basic ecosystem and value chain as shown in Figure 8 below.
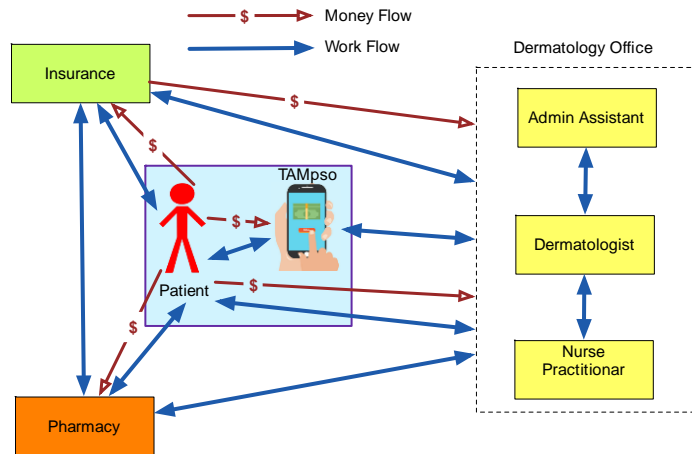
Figure 8. Ecosystem and Value Chain

The Figure 8 shows how the money flows between various stake holders of psoriasis treatment industry. Patient pays for insurance coverage to insurance companies, the dermatology office for the out-of-pocket or policy deductible costs, pharmacy for over-the-counter medications and prescription medications either not covered by the insurance or those that are not fully covered by the insurance. Insurance covers the part of the treatment costs and thus pays the dermatology clinic for the services rendered. TAMpso team initially believed that the revenue will be generated through payments from the psoriasis patients and the dermatologists for the services used.

In addition to the ecosystem, competitor analysis was carried out to understand the market space and is shown using the petal diagram in Figure 9. As shown in the petal diagram, there are 4 adjacent market segments: mobile apps, patient advocacy groups, dermatology groups, social networks. These are the markets where our future customers belong to today. Subsequent 42 interviews with the dermatologists, nurse practitioners, physician assistants, medical assistants, sales representatives of the dermatology pharmaceuticals and their medical science

liaisons provided us a lot of information to further refine out value propositions and other parts of the business model canvas.
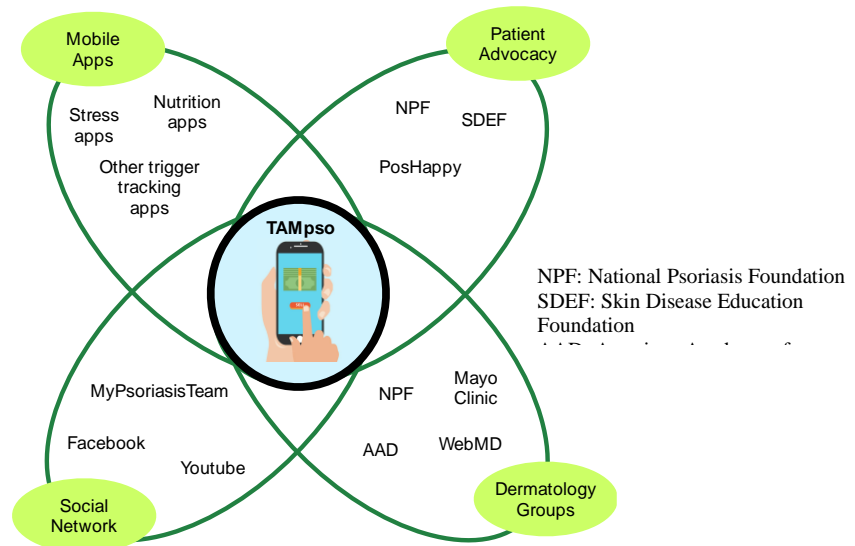
Figure 9. Competitor Analysis

First key information gained was that most of the dermatologist do not consider tracking triggers as very important for improving treatment. Therefore, this hypothesis was considered invalidated. Secondly, dermatologists rely on Body Surface Area (BSA) for assessment of psoriasis because it saves time as compared to Psoriasis Area Severity Index (PASI). What was surprising to know is that BSA documentation for patients is also required by the insurance companies. In addition, those who took images to support BSA documented in the clinical notes as evidence mentioned that the approval from insurance is more easily obtained with the images as supporting evidence. In addition, these images are shown to the patients on the follow-up visit to let them see the improvements in skin condition since their last visit to build confidence in the treatment. However, many of the interviewed dermatologists mentioned that

they do not acquire images due to time constraints and need to spend 5 to 10 minutes per patient for documenting the disease condition through clinical notes, which includes the BSA estimated through visual assessment, often resulting in inter- and intra-physician variability. This affects the efficacy of the treatment and is detrimental for controlling disease progression. Therefore, each patient goes through a trial-and-error phase to determine effective treatment, which can often take months to years. But dermatologists are slowly increasing the use of acquiring images for treatment documentation. Thirdly, clinical trials require PASI calculation because it is required by the US Food and Drug Authority (FDA) for approval of psoriasis treatment drugs. Newer medication based on PASI scores reported in the clinical trial are considered by the dermatologists to be more effective, but insurance companies usually does not cover the new medications. Therefore, insurance restrictions play an important role in determining the effective treatment plan for a patient. The reduction in time required and streamlining the process for paperwork associated with the clinical documentation required for the insurance approval for prescription medications for patients with moderate to severe psoriasis will be solving a major pain point for the dermatologists.

For the psoriasis patients, the key challenge is the treatment as patients want their skin to be clear from the lesions which most often does not happen as there exists no cure for psoriasis and the treatment only manages the symptoms. Patients find it hard to believe that a cure is not available despite the significant scientific advancements. As the patients start to realize the lifelong nature of the ailment, they tend to give up on treatment because of the fear of side effects, financial and family problems, forgetfulness, or lack of visible improvement over time,

resulting in problems in the patient–physician relationship. In fact, the prime reason for unsuccessful treatment is linked to discontinued therapy because of a lack of motivation in the patients. Patients lose confidence in the treatment and the care provider as response to medications often takes time, ranging from 3 to 6 months, during which time new flare-ups of psoriasis may also occur. In case of new flare-ups, it takes at least 3 weeks to see any noticeable changes with treatment. This makes it difficult for patients to notice the treatment response and keep track of disease progression solely through visual perception due to lack of severity assessment tools. When the patient perceives that the medications are not working, it leads to frustration and non-adherence. Manifestation of the disease causes social embarrassment to the patient as others treat it as a contagious disease, resulting in low social acceptance and depression.
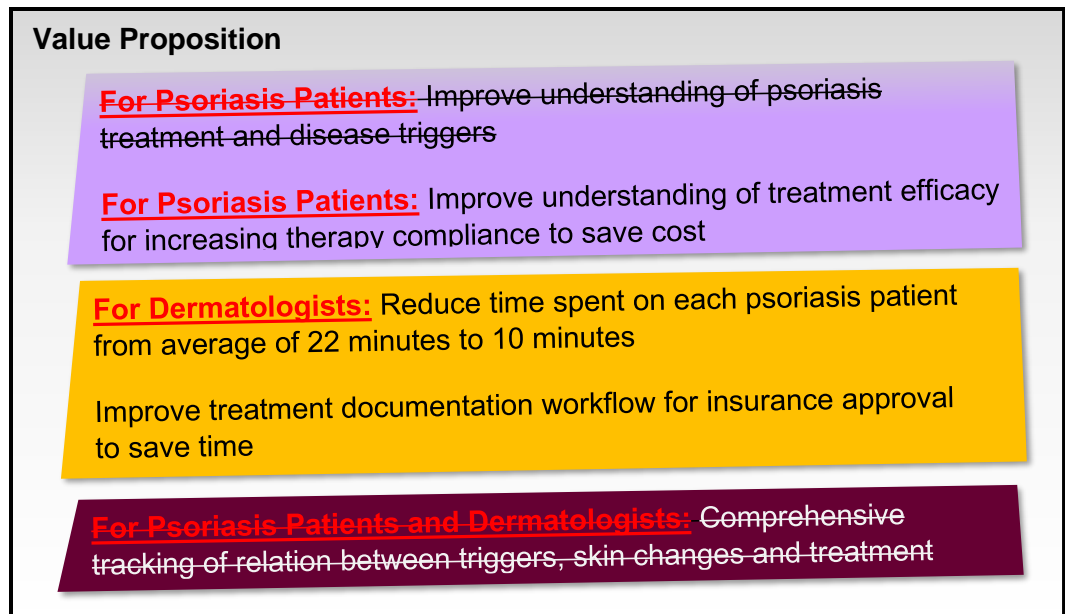
Figure 10. Value Proposition Revision 4

Lastly, the patient needs to keep record of all psoriasis treatments because if the provider is changed than ineffective treatments already tested might be repeated because of step therapy requirements set by the insurance. The revised value proposition based on the information above is shown in Figure 10. The interviews also revealed that the primary source of information for dermatologists regarding new developments in the field of dermatology is the Journal of AAD. The interviews with the sales representatives revealed that the companies either visit the dermatologist's office or visit the major events like AAD Annual Conference and Skin Disease Education Foundation (SDEF) conferences etc. for exhibiting their products to dermatologists. The initial channel diagram for reaching the customers is shown in Figure 11 below.
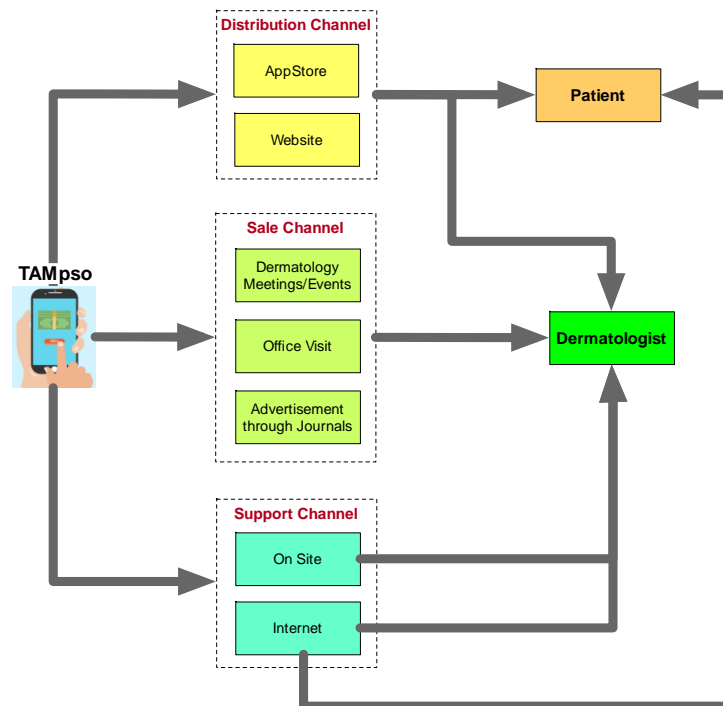
Figure 11. Channel Diagram

The Figure 11 provides the distribution channel, sales channel, and the support channel for the two customer segments dermatologists and psoriasis patients. Interviews with sales representatives have shown that dermatologists usually buy products through direct sales channel while for psoriasis patients web or tele-sales is more appropriate medium.

Another 11 interviews with dermatologist, patient advocate, pharmacy benefit management team representative, insurance (prior authorization) representative revealed that every medication has its own criteria for prior authorization which is a process through which approval from insurance is obtained for prescription medication. Each insurance plan sets its own criteria which depends on each patient's medication history. Therefore, maintaining medication history is critical in case of change of physician or insurance plan. Prior authorization is required for all biologics prescription usually prescribed to patients with moderate to severe psoriasis which may be valid for one fill or up to one year. Getting insurance approval for prescription medicine is a key challenge for the dermatologists. Clinical documentation of disease severity is required to prove the medical necessity for the prescribed medications for prior authorization which requires a lot of time-consuming paperwork. Even if providers know that initial medications may not be as effective for the treatment, providers still need to follow the requirements set by insurance to provide documented clinical confirmation.

The criteria for failure of treatment medication are minimal improvement observed in BSA over 3 to 6-month period. To obtain approval from insurance, physicians are required to provide a history of medication usage by the patients and documented failure of each

medication used. The documentation becomes more complicated if the patient has seen multiple providers, in which case the patient needs to have prior medication records indicating what he/she has been treated with in the past. Otherwise, the similar treatments might be repeated, and the patient may have to wait an additional 3 to 6 months before shifting to new medication. At times, providers may need to call prior treating physicians to ascertain the medications prescribed. Additionally, prior authorization may be denied even after submission of clinical documents because of a lack of coverage for the medication, in which case another medication needs to be prescribed, leading to more paperwork. The revised value propositions are provided below.

Value Proposition

For Psoriasis Patients: Improve understanding of treatment efficacy for increasing therapy compliance to save cost

For Dermatologists: Improve treatment documentation workflow for insurance approval to save time by 50% ($\approx$10 minutes) per patient
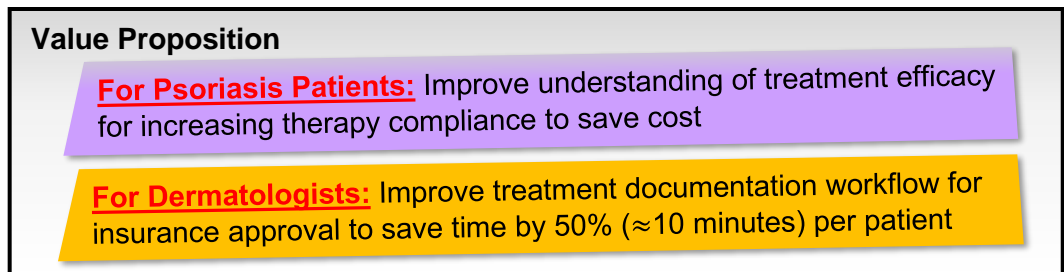
Figure 12. Value Proposition Revision 5

Based on the interviews conducted thus far and with the identification of sales channel and the important events for marketing information to the customer requires the estimation of budget. Therefore, the annual budget required for demand creation is provided in the table below.

Table 2: Demand Creation Budget and Forecast		
Activities	**Annual Budget**	**Forecast**
AppStore (Platform and Analytical Tools)	$250	Initial cost of $95 for iOS and $25 for Android. Analytical Tools will be used to see user activity within App (Flurry).

33

Website (Content Writer, Development)
   • Psoriasis Patients
   • Clinical Professionals
	$20,000	The professionally developed website will help convert customers from consideration to purchasers.

Conference Events
   • National Psoriasis Foundation (NPF) Events – Psoriasis Patients
   • American Academy of Dermatology (AAD) Events – Dermatologists
	Number of Events = 6
	Cost per Event = $5000
	Total: $30,000
		These events can help demonstrate the functionality of the technology to large number of people. Will help create awareness and interest by offering free live demonstrations.

Direct Sales • Full time salesperson salary and travel	$65,000	Dermatologist will not pay attention to product unless the representatives go in person to demonstrate.

Advertisements
   • NPF website and Facebook groups
   • AAD journal and website
	$15,000	National Psoriasis Foundation (NPF) website Ads and Facebook group will target Psoriasis Patients. AAD will create awareness among the dermatologists.

Budget Forecast	**≈ $130,000**	

The customer acquisition costs and the lifetime value for each customer is provided in the table below.

Table 3: Customer Acquisition Costs and Lifetime Value		
Customer	**Acquisition Cost**	**Lifetime Value**
Psoriasis Patients	$25,000 per 1200 customers $21 per customer	Average lifetime 3 years **$5 per month generated through subscription** 36 months based on average lifetime 1200 patients Total Value: **$216,000 (3 Years)** **Average $72,000 per year** **Lifetime Value per customer $180**

| Dermatologists | $105,000 per 25 customers

$4,200 per customer | Average lifetime 3 years

$5 per report generated as usage fee
1200 patients
3 reports per patient per year
25 Dermatologists
Total Value: **$1,350,000 (3 Years)**
Average   **$450,000 per year**
Lifetime Value per customer $54,000 |

As can be seen from the above table, based on our assumptions the Lifetime Value (LTV) per customer is much higher than the acquisition cost per customer. For psoriasis patients the LTV per customer is $180 compared to acquisition cost of $21 per customer. Similarly, for the dermatologists the LTV per customer is $54000 compared to acquisition cost of $4200 per customer. The estimated costs of channel incentives for get, keep, grow strategy are provided in table below.

Table 4: Channel Incentives and Cost	
Get Strategy	
Dermatologist	First 120 reports free ($600)
Psoriasis Patients	First 2 months free subscription ($10)
Keep Strategy	
Dermatologist loyalty (one year of use)	120 reports free ($600)
Psoriasis patient's loyalty (one year of use)	20% discount for two months ($2)
Grow Strategy	
Dermatologist referring another dermatologist	120 reports free ($600)
Dermatologist recommendation on website	20% discount for pre-authorization per patient for 120 reports ($120)

Dermatologist referring psoriasis patients	20% discount for pre-authorization per patient referred ($1)
Psoriasis patient's recommendation on website	20% discount on monthly subscription for one month ($1)
Psoriasis patients referring other patients	1-month free software service ($5)

Additional 15 interviews were conducted with psoriasis patients, clinic manager, pharmacy manager, patient financial service customer representative, insurance coordinator and nurse. The interview pointed out that the time saved from the documentation will enable physicians to see more patients for increased revenue and will provide more time for charting documentation and for the follow up phone call with the patients. Physicians are reimbursed for their services based on production work Relative Value Unit (RVU) which is standardized across the healthcare industry and physicians have the choice on deciding the number of RVUs they want to take per week. Based on the interviews with the patients, it was decided to drop the subscription costs for the patients, as patients may not be too inclined to pay for the service. Having a freemium service will encourage more users to avail the service which in turn can reduce the customer acquisition costs for the psoriasis patients. The revised customer acquisition costs, lifetime value and the revenue model are provided in the table below.

Table 5: Revised Customer Acquistion Costs and Lifetime Value		
Customer	**Acquisition Cost**	**Life Time Value (LTV)**
Psoriasis Patients	$5 per customer	
Dermatologists	$5000 per customer	Average lifetime 3 years. Reports per dermatologist per year 3000 Revenue per year per customer $15000 LTV per customer = $45000 Target: 10 Dermatologists per year

Revenue Model			
Customer	**Strategy**	**Revenue Stream**	**Tactics**
Psoriasis Patients	**Freemium**		**Freemium**
Dermatologists	**Usage Fee** Pricing per report generated. 3000 reports generated per year per dermatologist.	**Direct Sales** **Web sales** **Mobile App Sales**	**Value Pricing** **Volume Pricing** 20% discount on bulk purchase of 1000 reports

The value proposition was revised as provided below.

Figure 13. Value Proposition Revision 6

The workflow diagram was updated based on the interviews and is shown in Figure 14. As seen from the Figure 14, patient calls in for an appointment with the physician. Administration assistant or medical assistant obtains the patient information and schedules an appointment. Patient visits the dermatology clinic, provides insurance information to the administration assistant. The patient is then seen by the physician assistant who obtains the medication history, family history and prior illness details from the patient. Finally, the dermatologist sees the patient, diagnoses the condition, and prescribes medication. The patient goes to the pharmacy to get medication. The pharmacy checks with insurance, and if gets approval, the pharmacy will fill the prescription.
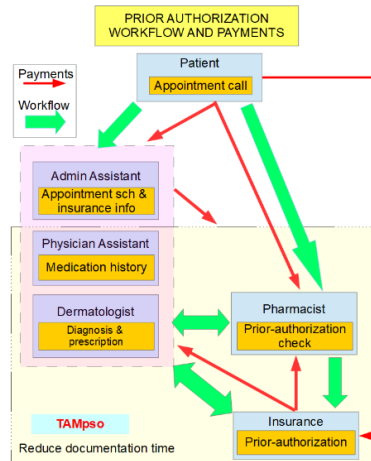
Figure 14. Payment and Workflow Diagram

If there is a prior-authorization requirement, the pharmacy lets the physician know, and the physician fills in the requisite forms. The form contains the physician and patient information, the treatment provided, and reasons for the medication prescribed and is faxed to insurance. The request may be approved within a day, or it may take up to a week. On average, 5 to 10 prior-authorization requests are made by the physicians per day which can vary from practice to practice. If prior authorization is approved, insurance sends it back to the physician where it is scanned and included with patient charts, and the pharmacy is told to fill the prescription. If the prior authorization is declined, insurance provides reasons for why it was denied. Based on the reasons, the dermatologist decides to either appeal the decision, in which case more documents are gathered to argue the decision, or she/he may decide to change the medication. If the appeal is done, then insurance can still deny it, and the provider will have to look for alternatives. At times, providers follow up with insurance to see if the requests are processed.

The table below lists the various roles of the customer archetype for the TAMpso technology.

Table 6: Customer Archetype				
Psoriasis Patient (User)	**Dermatologists (Buyer)**	**Medical Assistant (Recommender)**	**Administrative Assistant (Recommender)**	**Insurance (Influencer)**
Moderate to severe psoriasis >18 years old Visited dermatologist	Small independent practices ~ 10 Staff Prescribes medication Letter of medical necessity for prescription medicine	Obtain patient illness complaint and medication history Type of over the counter medication or home remedies already tried Symptom's information	Patient appointment scheduling Obtains patient personal information and insurance verification	Approves medication Decides the treatment approach

Additional interviews with the insurance claim representatives, pharmaceutical representatives, pharmacists, hospital chief financial officer, physician assistant and pharmacist pharmacy benefit management team revealed that there are two direct competitors or potential partners operating in the prior-authorization market space: 1) SureScripts CompletEPA and 2) CoverMyMeds. These competitors are slowly replacing the manual process of prior authorization with the electronic prior authorization, which integrate with existing EHR software. The electronic prior-authorization is a four-step automatic process which starts after the provider writes the prescription in EHR software. The provider is alerted to the medication requiring prior authorization. The provider completes requests and sends it to insurance. Insurance returns the decision. If approved, pharmacy fills the prescription. The problem with these electronic prior-authorization solutions is that the clinical information in which BSA is documented and the required justification for the medical necessity for the prescription is still written by the physician.

Based on this information the workflow was revised to improve the existing prior-authorization and offering a competitive substitute. The prior-authorization workflow starts with the patient entering personal and insurance information into TAMpso along with the prior medication history. Then the patient captures images of the psoriasis lesions, and the app determines the BSA, prepares a severity document, and shares it with the patient and the physician. When the patient visits the dermatologist, he/she has the severity documentation already available in the EHR software which determines the prescribed medication. If prior authorization is required for the medication, then the severity documentation along with patient insurance information is automatically sent to insurance. Insurance returns the decision. If approved, pharmacy fills the prescription. The workflow diagram is depicted in Figure 15. The partner box in the center "T" diagram demonstrates TAMpso features which require partnership with existing prior-authorization vendors for reducing the time to market the technology.

Interviews also pointed out that a software using existing Electronic Health Records (EHR) systems without affecting any structural changes have more chances of being adopted with quick approval times. There is a need for three-fold software integration between EHR software at the dermatology office (EPIC, Inc.), Insurance Pharmacy Benefit Management software (MedImpact, Inc.) and Pharmacy Management Software (PioneerRX, Inc.). The interviews with the pharmaceutical representatives let us know that the research and development scientists at pharmaceutical companies need to use a PASI score for assessment in clinical trials because of FDA requirements.
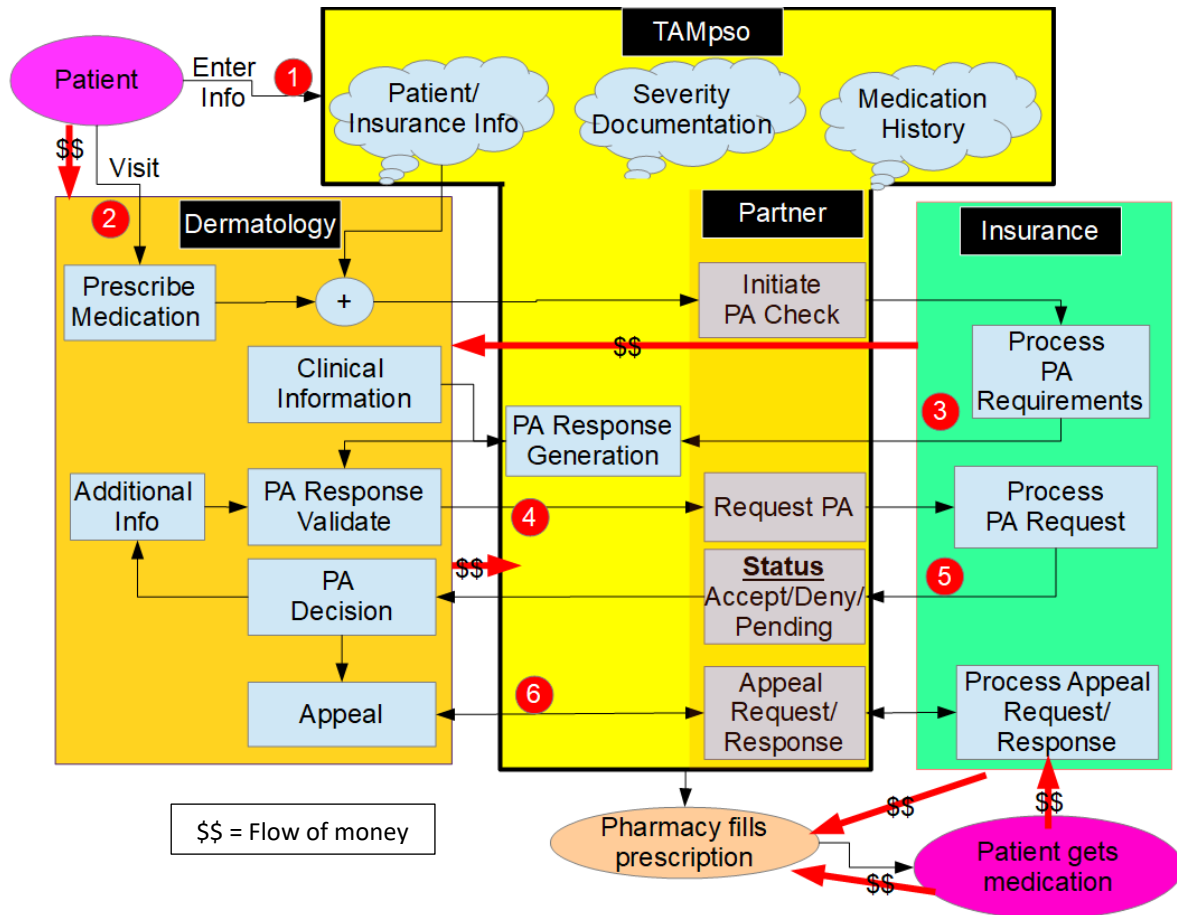
Figure 15. Prior-Authorization Workflow Diagram with TAMpso System.

But visual PASI evaluation suffers from both inter and intra-physician variability which requires intense training for the investigators at the start of the clinical trial. Additionally, to reduce it further requires increasing the number of patients to account for the variability or even a need to conduct multiple studies leading to an increase in cost for clinical trials. LEO Pharma (one of the major pharmaceutical companies active with psoriasis) and Roivant Sciences are actively searching for and trying to develop automated PASI solutions. Moreover, clinical trials usually compare study drugs with a placebo instead of other effective treatments,

resulting in a lack of information for comparing the efficacy of the various systemic and biologic treatments available for psoriasis. These interviews led us to identify and list potential partners as shown in the table below.

Table 7: Potential Partners			
Partner	**Costs**	**Risk**	**Benefits**
Existing Electronic Prior-Authorization Solutions: SureScripts, CoverMyMeds	Revenue Sharing	Intellectual property, Exclusivity	Can potentially reduce product to market time
Electronic Health Record (EHR) Providers Epic, Allscripts, CureMD, eClinicalWorks, GE Healthcare, Cerner, Practice Fusion, athenahealth, Kareo		Intellectual property, Exclusivity	Access to existing EHR databases of the independent practices is required to extract information
Dermatologist Private Practice	$10,000 per year		Help with execution and testing of the initial software development and deployment issues and generate user feedback
Pharmaceutical Companies Novartis, Celgene, Lilly, Pfizer			Establish brand name (Clinical trial for validating the automatic BSA calculation to increase acceptability)
UND Innovation Center	$35,000 per year		Office Suite, Fiber internet, Telephone, Access to copy, fax, scan, print, Staff/meeting room credit and utilities (heat, cooling, electricity)

Lastly, we were able to meet with the representatives of FotoFinder Systems, Inc. and saw detailed demonstration of automatic assessment of psoriasis disease using their latest clinical equipment. This helped us to understand how our technology is different from theirs and what competitive advantage we can offer. The final business model canvass as exported from launchpad central is shown in Figure 16 below.
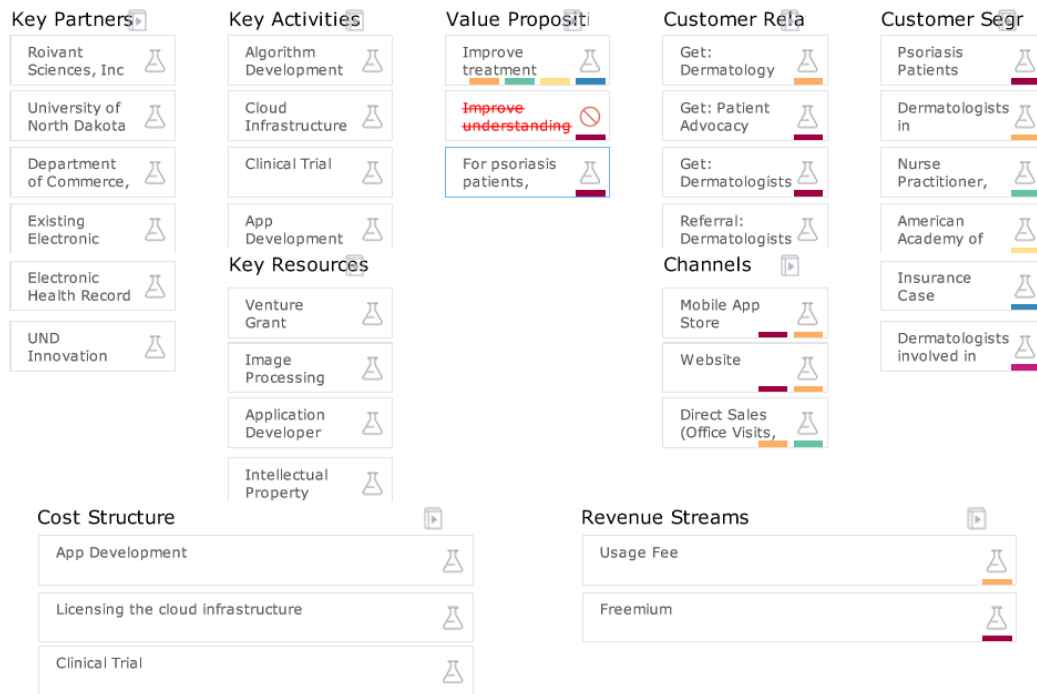
Figure 16. Business Model Canvass

Chapter Conclusion

The field interviews conducted as part of I-Corps program helped us to validate and correct most of our presumptions regarding the business model canvas. We were able to understand the complexity of healthcare sector within US and our interactions with various stakeholders has helped us to redefine some of our continuing research objectives.

MACHINE LEARNING BASED APPROACHES TO BIOMEDICAL IMAGE SEGMENTATION

In this chapter, an automatic segmentation and classification approach based on a three-step methodology for psoriasis images acquired using a digital camera is presented. Firstly, a k-means clustering based segmentation approach is applied that classify the images in normal and psoriasis lesion groups. Secondly, an approach for salient region detection is proposed which was used to quantify the level of erythema and scaling present inside the lesions. The approach was used to sub-divide the images into three groups: 1) red, 2) red + scaling, and 3) scaling group. Then, statistical analysis was used for color-based feature selection, corresponding to lesion characteristics within each sub-group. Lastly, the effectiveness of the selected features was validated using a support vector machine classifier and the effect of using variable pixel window sizes was evaluated.

Background

Machine vision has gained increasing importance in recent years as it provides the ability for imaging-based automatic inspection and analysis for wide ranging industrial applications. The range of machine vision applications in biomedical imaging is increasing due to innovative designs such as mini-cameras for endoscopy or minimally invasive surgery, imaging scanners to improve the quality of dentures and skin scanners to identify skin cancer among many others [21]–[23]. These innovations have resulted in the generation of an exponential amounts of the image data which has rendered manual analysis time and cost prohibitive and reduces the objectivity and reproducibility of the results. Hence, there is a growing need for automatic

image processing and analysis methods. The automatic image processing systems rely on the principles of the artificial intelligence (AI) which has enabled computers to do things which at present require strenuous human effort. While machine learning approaches has enabled machines to learn on its own without being explicitly programmed. Current recent examples of these include effective diabetes diagnosis system for improving the sensitivity and/or specificity of detection and diagnosis of disease which reduces the potential for human error in the decision-making process [24] and accurate diagnosis for malaria parasite detection and identification in stained blood smear images to remove the need for visual examination for more objective, time effective and error free diagnosis [25].

Like the above approaches, image processing can play an increasingly important role in objective assessment of skin disorders thus improving the management of chronic skin diseases since mostly these conditions can be diagnosed and assessed using clinically acquired digital images. One such chronic skin disease is known as psoriasis. There are many severity assessment techniques for psoriasis available in the literature [26], [27]. Among them Psoriasis Area Severity Index (PASI) and Body Surface Area (BSA) are the two most widely adopted techniques. PASI assigns scores to lesion severity based on three characteristics: erythema (total skin redness), induration (thickness of the scales) and scaling (amount of scales presents within the lesion) - for four body areas (head, torso, upper extremities, and lower extremities) weighted by the percentage of the body surface area (BSA) covered by the lesions [28]. PASI is the gold standard for assessment of psoriasis severity and is used in clinical trials [29]. However, the complexity of the score limits its use in clinical practice where BSA is more

prevalent due to its efficiency and convenience of use [30]. A new problem arises from the lack of correlation information that can translate the clinical trial findings based on PASI scores to BSA values used in daily practice along with generalization across different drugs [31]. The current dermatological guidelines for assessment of psoriasis suggest combining PASI score with the Body Surface Area (BSA) and Dermatology Life Quality Index (DLQI) which is validated questionnaire for determining the patient's quality of life [32]. This necessitates the need for an automatic approach to the objective analysis of psoriasis disease severity for evaluating the treatment efficacy [33], [34].

The widespread availability of smartphones has made it possible to remotely monitor psoriasis severity progression through mobile collected images with analysis on a cloud server and automatic reporting of assessment results to the patients and physicians, thus eliminating the need for personal clinical visits. The initial step for designing such a system requires implementation of suitable image-processing techniques, including segmentation that can separate psoriasis lesions from normal skin regions. Segmentation is necessary for estimating the BSA values and for the subsequent analysis of the lesion to characterize and estimate the erythema, induration, and scaling content present within the lesion.

The objective of the work presented in this chapter is to determine a set of color features that can be used to characterize the differences between psoriasis lesions from normal skin and accurately and automatically segment them. The approach used for psoriasis segmentation and feature selection is based on a previous work on *k*-means clustering-based approach for segmentation [35]. This work improves the existing segmentation technique along with an

analysis of statistically significant features for classification, and an analysis of the effect of variable window sizes for region-based psoriasis labeling is presented. In addition, a scaling contrast map-based approach is proposed for salient region detection for identification of the type of psoriasis images. The region detection is important as it determines the type of color features to be extracted for the analysis of the lesion. In the rest of the chapter, details of the image database, the k-means clustering approach, scaling contrast map, statistical analysis, and the Support Vector Machine (SVM) classifier are presented. At the end, the results and discussion are presented.

Image Database

For this research, the images were obtained from a publicly available database containing 45 psoriasis images along with the corresponding ground-truth images (psoriatic lesions in the image outlined by dermatologists) [36], [37]. The images were obtained using a 16-mega pixel 3X magnification digital camera in an uncontrolled setting under varying lighting conditions, distances, viewing angles, and backgrounds. Then, the images were manually cropped such that all the unwanted background was removed with only the skin and psoriasis regions remaining. The resolution of the cropped images varied between 154 by 151 to 1800 by 1700 pixels. Three representative images from psoriasis sub-groups; red, red + scaling and scaling are shown in Figure 17(a). The process for identifying the appropriate group for a given image is explained later. Manual annotation was performed with each pixel labeled as either healthy or psoriasis (ground-truth) and verified by expert dermatologists as shown in Figure 17(b).

Psoriasis Sub-Groups	Image Segmentation using Ground-Truth			
Red				
Scaling				
Red + Scaling				
	(a) Original Image	(b) Ground-Truth for Psoriasis Lesions	(c) Psoriasis Lesions	(d) Normal Skin
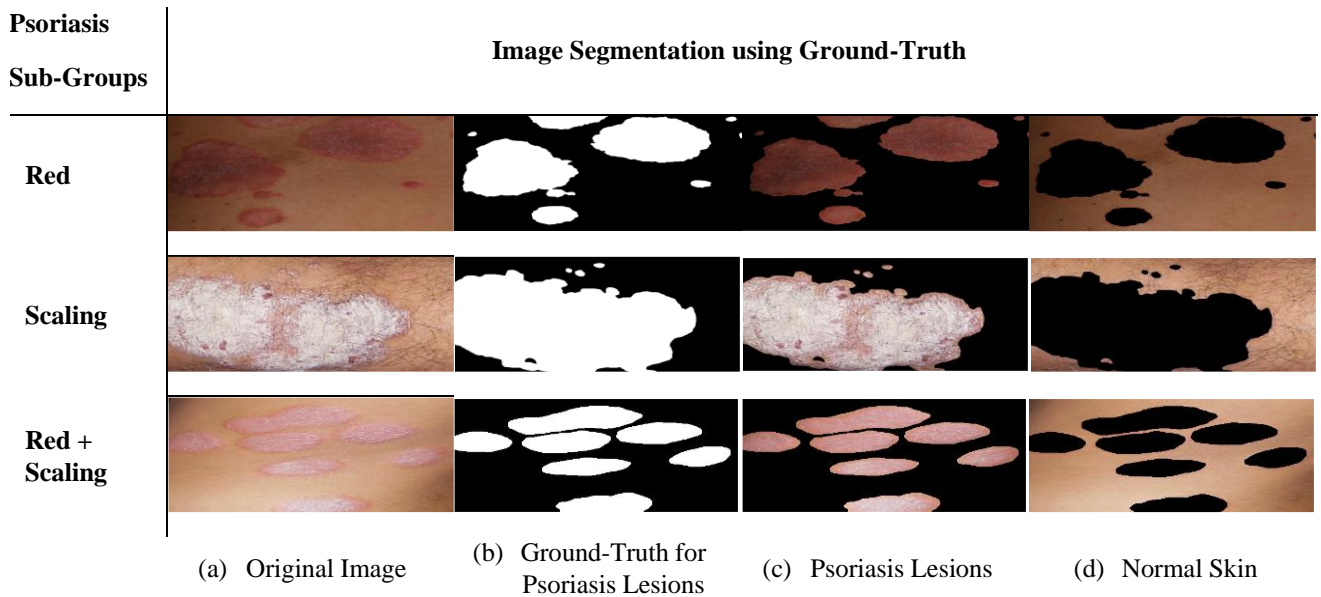
Figure 17. (a) Original images for the three sub-groups are shown. (b) Ground-truth image associated with the psoriasis lesions are shown. (c) Psoriasis lesion extracted using the ground-truth image and (d) Normal skin obtained using the inverse of the ground-truth image.

Figures 17(c) and (d) are the segmented images containing psoriasis lesion and normal skin respectively and are assumed to be 100% accurate.

Segmentation of Psoriasis Region

In medical imaging, segmentation refers to the division of an image into multiple regions (sets of pixels) such that each pixel in a given region shares common characteristics with other pixels in the region and are thus assigned a unique identifying label. The psoriasis segmentation techniques involve segmenting the psoriasis regions from the normal skin regions within the image. Most such techniques described in the literature involve color, texture and active-contour based approaches [38]–[43]. Texture-based methods require parametric optimization for various texture regions which are useful in certain scenarios but usually difficult to generalize [40]. The contour based techniques are noise sensitive, depend on the initial contour selection and mostly applicable to images with uniform sub-regions but without internal edges [39], [44]. The variation in skin color associated with psoriasis lesions such as silvery white color due to lamellar scales or redness of the skin due to erythema has led researchers to use

color features for extracting valuable information to distinguish between psoriatic lesions and healthy skin. Existing literature studies have involved analysis of exhaustive color feature space consisting of the 14 known color spaces [45]. These studies have either focused on the analysis of discriminating characteristics of the color features tested on manually segmented skin and psoriasis images [45] or have considered pixel-based approaches [46]. The former approach does not specify how to automatically segment psoriasis and skin regions while the latter is time consuming and computationally expensive as it involves extracting multiple features per pixel of the image. A k-means clustering with morphological processing technique for psoriasis tracking based on red-green-blue (RGB) color-skin images was proposed in [47]. The problem with that approach is the RGB color space is affected by the illumination intensity which varies depending upon the location of the image capture and thus the performance degrades [48], [49].

We applied the k-means clustering approach which clusters the image pixels into two groups; psoriasis lesions and the normal skin; based on the similarity of the color characteristics [35]. The approach used was based on the algorithm developed by Arthur et al. [50] in which the image is divided into k mutually exclusive partitions based on the squared Euclidian distance between a cluster center and an image pixel. The idea is to find a space in which pixels within each cluster are as close together as possible based on the selected distance metric and maximally distant from pixels in the other clusters. Prior to the clustering, the images were first converted to the device independent, International Commission on Illumination (CIE) Luminance and color-components (L*a*b*) color scale and the luminance information was

separated from the color information. The luminance component (L*) was removed as it varies depending on the illumination conditions at the time of image capture and is rarely consistent. The color components (a*b*) were used for the k–means clustering operation with $k = 2$ clusters (psoriasis and normal skin). k-means clustering is an iterative process of determining k centroids such that the total sum of the distances between each image pixel within the cluster is the global minimum for the determined centroid. In practice, to avoid local minima with lower total sum of distances compared to the computed centroids the process is replicated several times with random starting points. In our experiments, the process was replicated five times. The clustering operation divided the image pixels into two different clusters labeled as either psoriasis lesion or normal skin.

After clustering, the segmentation was improved using morphological operations. The morphological operations were performed to remove holes within the segmented region using morphological reconstruction algorithm described in [51] followed by dilation and then erosion. A linear structuring element was chosen for the dilation operation, and a diamond structuring element was used for the erosion operation. After the morphological operations, a guided filter was used, as described in [52], to retain the edges for extracting the border of the segmented region. After segmentation, the segmented psoriasis regions were converted to binary images and a segmentation mask was obtained. The quantitative comparison was performed using the dice coefficient to measure the similarity with the binary mask of the ground-truth images [53]. Dice's coefficient is defined as

$$P(A,B) = \frac{2|A \cap B|}{(|A| + |B|)} \times 100\,\% \tag{1}$$

where *A* and *B* show pixels in the binary ground-truth image and the binary segmented image, respectively.

Scaling Contrast Map

The *a** dimension of the L*a*b* color space represents the red–green color component with positive and negative values indicative of red and green color respectively. The color of erythema (redness) corresponds to positive values of *a** and its inversion enhances the contrast between erythema and the adjacent skin or scaling. Therefore, in order to enhance the contrast, a scaling contrast map proposed in [41], [54] is modified as

$$S_{x,y} = J\left(inv\left(a^*_{x,y}\right)\right) \tag{2}$$

where $S_{x,y}$ is the contrast enhanced pixel value at the image coordinate *(x,y)*, $J(.)$ is the multi-scale center surround filter and $inv\left(a^*_{x,y}\right)$ inverts the *a** dimension as

$$inv\left(a^*_{x,y}\right) = \max_{i,j}(a^*_{i,j}) - a^*_{x,y}, \begin{array}{l} i = 1,2,\dots,M \\ j = 1,2,\dots,N \end{array} \tag{3}$$

where M and N are the total number of pixel rows and columns. The multi-scale center-surround filter $J(.)$ is defined by:

$$J(X_{x,y}) = \sum_{s=1}^{3}\left(X_{x,y} - \frac{1}{k} \sum_{\substack{x-w(S)\leq m \leq x+w(s) \\ y-w(s)\leq n \leq y+w(s)}} X_{m,n} \right) \tag{4}$$

where s is the scale. The scale determines the window size w, $w = d/2^s$, where d is the larger pixel dimension of the image width or height, and k is the number of pixels in the window. The variable window size corresponding to different scale sizes allows the detection of contrast both locally and globally for a given image pixel.

For each image, at the end of the k-means clustering segmentation, two masks are obtained, one corresponding to psoriasis lesion and the other for the normal skin. These masks are used to segment the lesion region and the normal skin region in the contrast enhanced image. Then the mean values of the two segmented images are calculated to determine the average inverted contrast enhanced a^* value corresponding to the lesion and the skin region. Following equation is proposed as a measure of quantification of the presence of erythema and the scaling within the psoriasis digital images:

$$T = \frac{L}{L + NS} \tag{5}$$

where L is the mean value of the lesion region and NS is the mean value of the normal skin region. By analyzing the equation, we can see that for an image containing no lesions, $T = 0$, meaning the image only contains the normal skin. The image containing only the lesion will result in $T = 1$. The values between 0 and 1 varies depending upon the presence of the erythema and the scales within the lesion. Higher erythema content results in lower values of $T \sim 0$, while higher scaling content results in higher values of $T \sim 1$. Since the formation of a psoriasis lesion changes the skin color to reddish (erythema), silvery (scaling) or a combination of both

52

(erythema and scaling), we subcategorized the 45 images into the three groups based on the value of T using criteria,

$$Group = \begin{cases} Red & 0 < T \leq 0.25 \\ Red + Scaling & 0.25 < T \leq 0.5 \\ Scaling & 0.5 < T \leq 1 \end{cases} \tag{6}$$

Initially the images were sub-divided in to the three groups through visual inspection. Then using Eq. 5, T values and threshold ranges were determined heuristically, as provided in Eq. 6, to obtain image groupings that were consistent with the visual inspection results. The resultant number of images within each sub-group are provided in Table 8 and representative images of each sub-group is shown in Figure 17.

Table 8: Psoriasis Sub-Groups		
Sub-Groups	**Psoriasis Region Properties**	**Number of Images**
Red	Erythema only	8
Scaling	Scaling only	16
Red + Scaling	Erythema and Scaling	21

Statistical Analysis of Color Features

We hypothesized that a set of robust color features could accurately distinguish between psoriasis lesions and normal skin. To validate this, statistical analysis was performed for color features extracted from various color spaces. Each image within the three sub-groups was divided into two images corresponding to psoriasis and skin regions using the ground-truth mask as shown in Figure 17(c) and (d). The mean value characterizing the average color corresponding to the following individual color components was evaluated for the two extracted images after color space conversion:

a) R, G, and B channels in red-green-blue (RGB) color space,

b) *Y*, *Cb*, and *Cr* channels in luminance and chrominance (*YCbCr*) color space,

c) *H*, *S*, and *V* channels in hue, saturation and value (*HSV*) color space, and

d) *L**, *a**, and *b** channels in lightness and color-component (*CIE 1976 L*a*b**) color space.

The 24 mean feature values, 12 each for psoriasis and skin regions, were determined for all the 45 images in the dataset and statistical analysis was performed to identify which of these features provided statistically significant differences between the normal skin and psoriasis regions for the critical value of 5% as the test of significance. For statistical analysis, the normality of data was first evaluated using the Shapiro-Wilk test [55]. Based on the normality, either the independent sample two-tail t-test or the Wilcoxon rank sum test was applied for the analysis using the 'Statistical' toolbox in Mathworks MATLAB software [56]. The statistically significant color features corresponding to each image sub-group (red, scaling, red + scaling) were determined.

Classification using Support Vector Machine (SVM)

After the statistical analysis, the two images obtained from segmentation were divided into non-overlapping 16-by-16 pixel regions. Table 9 shows the total number of images within each sub-group and the corresponding number of 16-by-16 non-overlapping windows obtained.
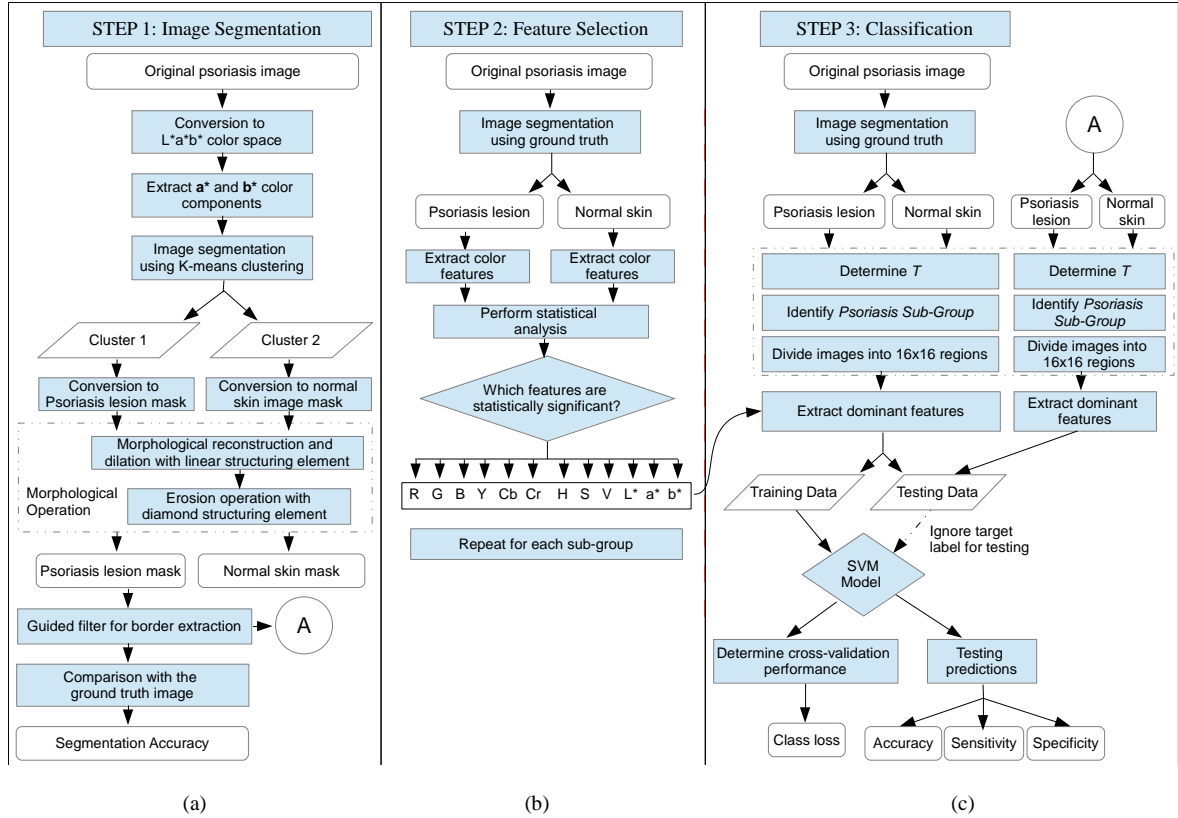
Figure 18. Methodology of the proposed system is shown. (a) The first block shows the approach followed for image segmentation. (b) The second block shows the approach for feature selection to determine statistically significant or dominant color features for each sub-group. (c) The third block shows the classification approach followed for distinuishing between psoriasis regions and the skin regions.

Table 9: Number of Training and Testing Windows In Psoriasis Sub-Groups				
Sub-Group	Number of Images	Total Windows (16x16)	Training Windows	Testing Windows
Red	8	686	460	226
Scaling	16	1732	1158	574
Red + Scaling	21	1812	1598	214
Total	45	4230	3216	1014

As shown in Table 9, the red + scaling sub-group has 5 more images than the scaling sub-group, but the number of 16-by-16 windows are slightly more than the scaling sub-group. This is due to the difference in the resolutions of the images in each sub-group. For each sub-group, statistically significant color features were evaluated and used to train and test a Support Vector

Machine (SVM) classifier with a linear kernel function to classify the 16-by-16 pixel regions into diseased and non-diseased classes. The number of training and testing windows are provided in Table 9. Figure 18 shows the entire methodology.

Figure 18(a) shows the *k*-means clustering based segmentation approach that is used to segment the digital image into normal skin and psoriasis lesion image. These images are compared with the ground-truth images to determine the segmentation accuracy. Figure 18(b) shows the statistical analysis approach for the determination of statistically significant color features. Each image is first segmented into psoriasis lesion and normal skin image using the ground-truth image. For each of the segmented images the 12 color features are determined for all 45 images. Statistical analysis is performed to analyze the difference for each color feature to determine statistical significance within the three psoriasis sub-groups determined using Eq. 5 and 6. Based on the analysis, the statistically significant features which are the dominant features are selected for each psoriasis sub-group. Dominant features for each psoriasis sub-group form the feature vector for the SVM classifier. The non-significant features are not considered for the subsequent analysis. In Figure 18(c), the images are segmented using the ground-truth images. The ground-truth segmented images and *k*-means segmented images are divided into sub-windows of 16-by-16 window size. The 16-by-16 window size was determined to be the optimal size through analysis with variable window sizes. The details are provided later in the results section. These windows are divided into training and testing data as provided in Table 9. For training data, 10-fold cross validation was used for evaluating the classifier learning performance by determining the classification loss or class loss. For testing data, the

results are obtained for ground-truth segmented images and *k*-means segmented images. The ground-truth segmented images provide the reference maximum accuracy obtainable with the SVM classifier for classification of skin and psoriasis lesions. Following performance parameters were evaluated for the SVM classifier:

a) Classifier accuracy was determined using the ratio of correctly classified testing windows to total number of testing windows classified. The testing windows comprised of both the psoriasis lesion and the normal skin.

b) Sensitivity of the classifier was evaluated as the ratio of correctly classified psoriasis lesion windows to total number of lesion windows.

c) Specificity of the classifier was evaluated as the ratio of correctly classified normal skin windows to total number of normal skin windows.

For the evaluation of the performance parameters, 'Bioinformatics Toolbox' in MATLAB was used [57].

Results And Discussion

This section provides the outcomes of the above three analyses. First, we obtained the results of the scaling contrast map and identification of psoriasis sub-groups and statistical analysis of the 12 color features for three sub-groups: red, scaling, and red + scaling. Secondly, we analyzed the segmentation accuracy achieved using *k*-means clustering approach for all 45 images using the available ground-truth masks. Lastly, for the third step, the average accuracy, specificity, sensitivity, and class loss were calculated to analyze the performance of the SVM classifier.

Results of Scaling Contrast Map and Statistical Analysis

The results of region identification using scaling contrast map are shown in Figure 19(a). and statistically significant color features corresponding to each of the identified image sub-groups (red, scaling, red + scaling) are shown in Figure 19(b). As shown from Figure 19(a)., for the scaling sub-group the absence of erythema results in the mean value of the lesion region to be higher and either closer to the mean value of the normal skin or higher than the normal skin resulting in T-value from Eq. 5 to be 0.5 and higher. $T3$ is the threshold value between scaling and red + scaling sub-group. On the other hand, for the red sub-group, the presence of erythema, lowers the mean value of lesion region, as erythema appears as black in the lesion after contrast enhancement and therefore has lower inverted a* values consequently T-values are taken as lower than 0.25. $T1$ is the threshold value between red and red + scaling sub-group. The red + scaling group has intermediate mean values depending upon the level of presence of erythema and scaling within the lesion. One of the problems with the scaling contrast map was high computational time and to reduce it images were downsized to 30% of their original size. The comparison of T-values for the original image and RS-value corresponding to the resized image indicate that for high resolution images the images can be downsized without any change in result thus solving the problem of high computational time.

Psoriasis Sub-Grouping
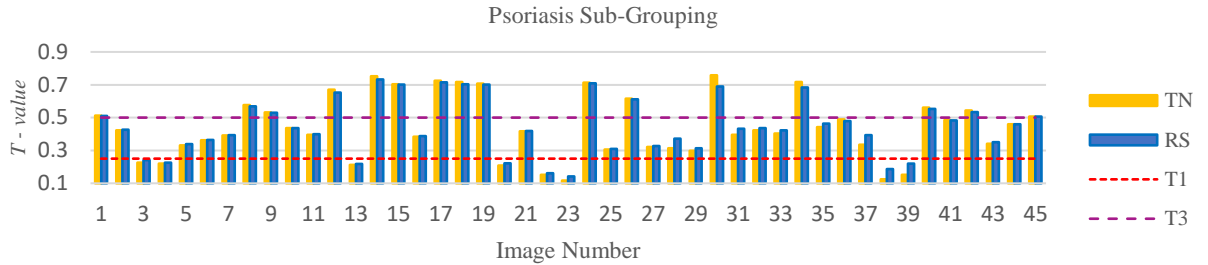
TN = Raw Image T-Value, RS = Resized Image T-Value (Downsized by 30%)

Figure 19(a). Psoriasis sub-group identification for images using modified scaling contrast map. TN and RS represents the T-values corresponding to each of the original and downsized images respectively. T1 is the threshold separating red and red + scaling sub-group while T3 is the threshold separating red + scaling and scaling sub-group.
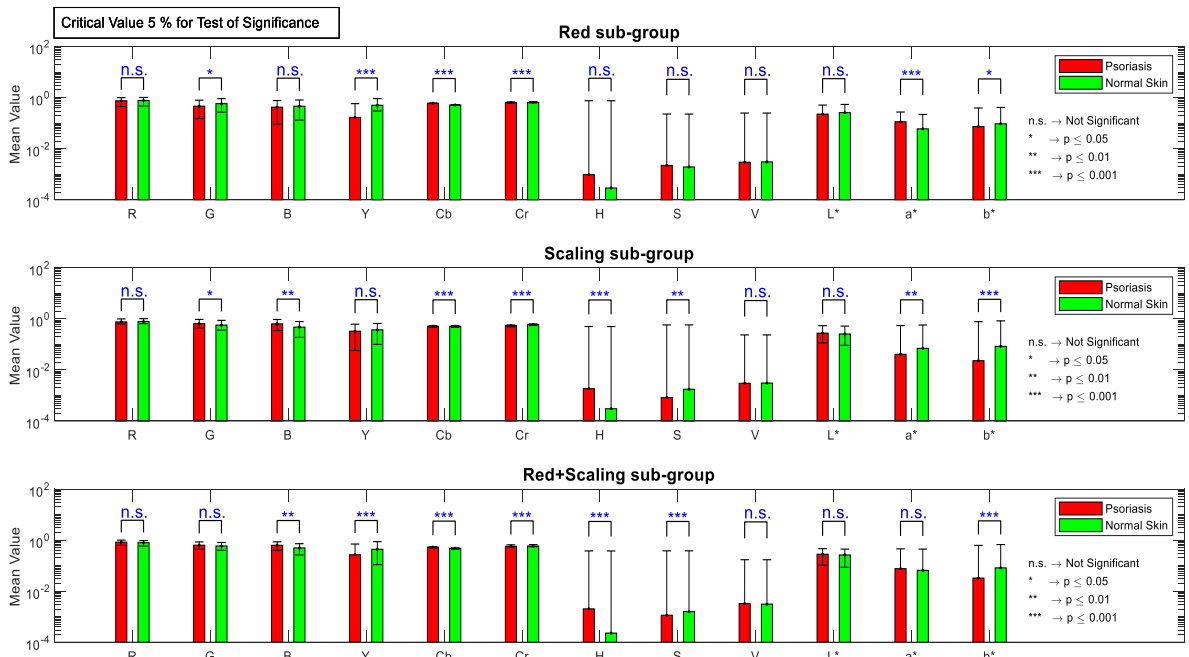
Figure 19(b). Results for test of significance for differences between normal skin and psoriasis lesion for all color channels. Blue stars specify the range of p value (logarithmic scale on y-axis) and the level of statistically significant difference. The red bars indicate the mean color value of lesions whereas green bars indicate the mean value of normal skin. Error bars indicate the level of variation (minimum and maximum values) from the mean values for all images in the sub-group.

Statistical analysis revealed that for the red sub-group the mean value of the green (G), luminance (Y), blue chroma (Cb), red chroma (Cr), a* and b* color components have statistically significant differences between the normal skin and psoriasis regions. Based on this analysis, for the red sub-group of images the input feature vector for the SVM classifier is comprised of only these six features. Similarly, for the scaling sub-group and the red + scaling sub-group the feature vector is comprised of 8 and 7 statistically significant feature values respectively as shown in Figure 19(b). The analysis also revealed that chroma blue, chroma red and b* color component had statistically significant differences for all three sub-groups. The b* color component falls along the blue–yellow axis. The mean value of b* component for psoriasis lesion was found to be less than the mean value for the skin region.

The reason is that normal skin usually has less red and more yellow color, and thus a higher b* value whereas the psoriasis lesions have higher red component and a lower b* value. It should be noted that hue and saturation have near zero mean values for the scaling and red + scaling sub-groups. The reason is that saturation is expressed in terms of the absence of white color. For these two groups the overall color is dominantly white due to the presence of silvery white scales as shown in Figure 17(a) resultantly the saturation is near zero. For the red + scaling group, the red color associated with the erythema and white color due to the scales cause the skin color to appear as pink as seen in Figure 17(a). The value or brightness component of HSV color space and L* the lightness or luminance component of L*a*b* color space was found to be non-significant for all sub-groups which indicates that lightness or darkness of the lesions does not constitute significant differences for differentiating between

psoriasis lesion and normal skin. The reason is that the brightness or intensity of pixels is almost the same across the image and is dependent on the external lighting conditions. Another important point to note is that the red color component of the RGB color space was not significant across all sub-groups even though for psoriasis we are interested in estimating the level of redness (erythema) as one of the parameters of PASI score. The reason for that as mentioned earlier is that psoriasis lesion does not have pure red color either due to presence of scales or due to underlying skin tone which may be anywhere from dark to light in color. For this reason, *Cb* and *Cr* color components are much better choice as shown in Figure 17(b) to have high statistical significance with $p < 0.001$.

Results of k-means Clustering

The results of the *k*-means clustering approach are shown in Figure 20. The first row of the Figure shows a sample of the original images in red, scaling, and red + scaling categories and the second row displays the scatter plot outputs of the *k*-means clustering algorithm applied on the a* and b* components of the original images for all three sub-groups. As seen from the scatter plot (Figure 20, second row), the psoriasis region and normal skin region have been successfully partitioned into two clusters. The output clusters of the *k*-means are converted back into the image masks as shown in the third row of Figure 20. The outlines of the resulting mask cover all the psoriasis regions but leave several holes. To fill the holes, morphological operations are performed. The results of the red and scaling sub-groups match the ground-truth with accuracy $> 98\%$, but for the red + scaling sub-group the morphological operation resulted in the removal of some of the psoriasis lesions.

Sub-Groups	Red	Scaling	Red+Scaling
Original Image			
k-means clustering			
Segmentation outcome			
Morphological Operation and Guided Filter Outcome			
Lesion Border			
Segmentation Accuracy	99.01%	98.40%	91.35%
Dice Similarity Coefficient	94.92%	98.61%	87.60%

Figure 20. Results of the proposed k-means clustering-based segmentation. First row contains the original psoriasis images. Second row contains the scatter plot showing the results of k-means clustering and the centroid locations. Third row shows the segmentation results obtained through k-means. Fourth row shows the result of morphological operations and the guided filter. Fifth row shows the final lesion border obtained through proposed segmentation technique. Last row provides the accuracy achieved for each of the images.

This is due to the difference in image resolutions which are 1100 by 900, 850 by 650, and 306 by 303 for red, scaling, and red + scaling sub-groups examples, respectively. The size of the structuring element used in the morphological operation was fixed for all three images resulting in the removal of small psoriasis regions in the lower resolution images. The average accuracy of the *k*-means clustering based segmentation for all 45 images was found to be $\approx$ 92% which is around 10% higher than the existing reported results as given in Table 10.

Table 10: Comparison of Psoriasis Segmentation Accuracy on the Same Database	
Methods	**Segmentation Accuracy**
Fuzzy C-Means (FCM) Clustering [39]	77.79%
Fuzzy Kernel C-Means (FKCM) Clustering [39]	78.80%
Semi-Wrapped Gaussian Mixture Model (SWGMM) Clustering [39]	82.84%
Proposed K-Means Clustering with Color Feature	92.12%

The average value of dice similarity coefficient for the *k*-means clustering based segmentation for all 45 images was found to be 86.87%.

Results of SVM Classification

The pixel-wise labeling of psoriasis lesions and normal skin is computationally expensive as it requires multiple features to be evaluated for each pixel of the image. Therefore, we analyzed the accuracy of the SVM classifier for various window sizes to identify the optimal window size and the results are provided in Table 11.

Table 11: SVM Performance for Varying Window Size					
Sub-Groups	**Number of Test Images (ω_i)**	**Window Size**			
		4x4	**8x8**	**16x16**	**32x32**
Red	3	95.68	97.59	99.07	98.21
Scaling	5	68.82	72.61	68.26	68.37
Red + Scaling	6	86.61	84.42	90.17	89.04
Weighted Average (%)		82.20	83.02	**84.25**	83.62

The 16-by-16 window size was found to be optimal for all regions as it achieved the highest

weighted average accuracy of 84.25%. The weighted average was calculated using the equation,

$$Weighted\ Average = \frac{\sum_{i=0}^{N} \omega_i x_i}{\sum_{i=0}^{N} \omega_i} \tag{7}$$

where ω_i is the individual weight and is equal to the number of test images and x_i is the accuracy

of the system within a sub-group category. The test images were randomly selected from each

sub-group and the process was repeated 10 times, each time with a different set of training and

test images. The results provided in Table 11 are the average results of the 10 trials.

During the training of the SVM classifier 10-fold cross validation was used to determine

the class loss for in-fold observations using the model trained on out-of-fold observations to

determine the predictive inaccuracy of the trained model. The class loss during classifier

training and the sensitivity and specificity results for the test images of the SVM classifier are

provided in Table 12. These results represent the average for the 10 trials.

Table 12: SVM Classifier Performance for 16x16 Windows for Ground-Truth Images			
Sub-Groups	**Sensitivity (%)**	**Specificity (%)**	**Class Loss (%)**
Red	98.13	100	0.11
Scaling	67.81	68.86	21.99
Red + Scaling	92.99	87.36	6.85
Weighted Average (%)	**85.10**	**83.46**	**10.81**

The class loss, which is the estimated out-of-sample misclassification rate, of 0.11% and 6.85%

for the red and red + scaling group indicate that the trained SVM can generalize and has a lower

predictive inaccuracy. This can be seen from the accuracies of the red and red + scaling sub-

groups provided in Table 11 which are 99.07% and 90.17%, respectively for the 16-by-16

windows. On the other hand, the class loss for the scaling sub-group is 21.99%, a high value

indicating that the trained predictor model has higher predictive inaccuracies. This reflects the classifier accuracy of 68.26% for the scaling sub-group shown in Table 11. One way to improve the classifier accuracy is to change the kernel function for the SVM from linear to polynomial of higher order. However, this would result in a higher chance that the model will overfit, causing the predictive inaccuracy to be greater even though we may get lower class loss.

Table 13: SVM Classifier Performance for k-Means Segmented Images			
Sub-Groups	Sensitivity (%)	Specificity (%)	Accuracy (%)
Red	89.53	85.32	87.43
Scaling	58.01	86.17	72.09
Red + Scaling	94.41	68.76	81.59
Weighted Average (%)	80.36	78.53	79.45

Table 13 provides the results for SVM classifier for k-means segmented images. As shown the accuracy is slightly lower than the results for the ground-truth images in Table 12. The classifier accuracies for red and red + scaling sub-groups validate the statistical analysis carried out for the color-based feature selection for differentiating psoriasis and normal skin regions. Our window-based approach reduces computational complexity and achieves higher average segmentation accuracy as compared to pixel-based approach. A pixel based approach using SVM classifier, reported in [39], achieved an average classification accuracy of 63.56% for the 40 images from the same image database that is used in this study. These results are considerably lower than what we achieved through our algorithm. An approach for skin hair removal was proposed in [58] for pre-processing of the psoriasis images. It was reported that presence of hair in the psoriasis images adversely affects the feature extraction thus increasing the misclassification rate. A similar preprocessing approach for hair removal was used by [39]

as well. Our approach does not require any pre-processing for hair removal thus making it more robust than existing approaches.

In all the results, the scaling sub-group has the lowest performance which makes sense as the color of the lesion because of the dominant presence of the scales is mostly white and often very similar to the color of the skin. Also, in some images, the flashlight of the camera was found to be one of the factors contributing to the lower accuracy as the reflection of the light on portions of the normal skin had the similar appearance as scales in the digital images.

Fuzzy C-Means (FCM) and Gaussian Mixture Model (GMM)

In k-means algorithm, the image pixels belong to only one cluster. Assuming there are k clusters than each pixel will be a part of only one cluster and cannot belong to any other cluster. The problem with this approach is that k-means will fail if pixels are part of overlapping clusters. To overcome this problem fuzzy c-means clustering algorithm was used in which each pixel, in addition to be a part of a cluster, is also assigned a membership weight which determines the degree to which a given pixel is considered a part of given cluster. This means that a pixel can be present in multiple clusters with varying degree of membership and is based on the minimal reduction of objective function,

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{K} u_{ij}^m \| x_i - c_j \|^2, \quad 1 \le m < \infty \tag{8}$$

where m is known as the fuzzification coefficient and is a real number that can have any value greater than 1. Its value is used to control the degree of fuzzy overlap with higher values

resulting in greater overlap between clusters. This results in increase in data points belonging to more than one cluster. N is the total number of data points, K is the number of clusters, u_{ij} is the degree of membership of ith data point to j^{th} cluster, x_i is the ith data point, c_j is the cluster center of jth cluster, $\|*\|$ is the norm function that expresses the similarity between the data points and the cluster centers. The similarity measure can be based on distance, connectivity, and intensity.

Fuzzy partitioning starts with random initialization of cluster membership values u_{ij} and then cluster centers are determined using the equation,

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m x_i}{\sum_{i=1}^{N} u_{ij}^m} \tag{9}$$

Then the clusters membership values u_{ij} are updated based on the equation,

$$u_{ij} = \frac{1}{\sum_{s=1}^{N} \left(\frac{\|x_i - c_j\|}{\|x_i - c_s\|} \right)^{\frac{2}{m-1}}} \tag{10}$$

Based on these determined values, objective function, J_m is calculated using Eq. (8). The process is iteratively repeated for optimization of the objective function by converging to a local minimum and is terminated either through reaching the pre-determined maximum number of iterations or if J_m reaches the desired level of optimization determined through the equation,

$$max_{ij}\{|u_{ij}^{s+1} - u_{ij}^s|\} < \varepsilon \tag{11}$$

where s is the iteration steps and ε is the pre-determined termination threshold with value between 0 and 1.

The GMM approach, on the other hand, is a generalization of k-means algorithm which considers the covariance information of the data to find a mixture of gaussian probability distributions that best fits the input data. The methodology as shown earlier in Figure 18(a) was adopted for evaluating both the clustering approaches.

The Table 14 below provides the segmentation accuracy for each of the psoriasis images in the database for k-means clustering, fuzzy c-means clustering and gaussian mixture model-based clustering algorithms. The green color in the table indicates an increase in accuracy in comparison to k-means clustering. Similarly, red color indicates a decrease in accuracy and no color indicates that there was no change in accuracy with respect to k-means clustering.

Table 14: Segmentation Accuracy			
Image	k-Means Clustering Algorithm	Fuzzy C-Means Clustering Algorithm	Gaussian Mixture Model based Clustering Algorithm
1	90.42%	90.48%	61.91%
2	92.16%	91.88%	93.90%
3	97.83%	97.83%	97.91%
4	99.01%	99.01%	96.47%
5	81.09%	80.87%	91.96%
6	98.24%	98.21%	97.69%
7	97.62%	97.62%	97.28%
8	87.44%	87.40%	82.03%
9	74.47%	73.80%	21.79%
10	72.65%	76.28%	86.67%
11	98.16%	98.10%	85.69%
12	98.00%	97.89%	96.03%
13	98.46%	98.49%	97.03%
14	95.09%	95.09%	94.00%
15	98.40%	98.40%	98.66%
16	98.74%	98.60%	86.00%
17	98.20%	98.20%	97.74%
18	96.06%	96.04%	90.37%

19	97.72%	97.63%	94.91%
20	70.16%	32.20%	98.07%
21	98.39%	98.09%	98.35%
22	95.84%	40.21%	95.00%
23	98.52%	96.94%	96.60%
24	90.85%	90.67%	88.87%
25	94.91%	94.87%	95.99%
26	95.27%	95.25%	88.06%
27	95.07%	93.16%	76.15%
28	87.17%	87.17%	87.83%
29	88.95%	89.29%	93.30%
30	92.55%	92.64%	92.48%
31	91.35%	91.99%	82.18%
32	93.17%	93.43%	92.72%
33	97.13%	97.43%	97.86%
34	88.83%	88.98%	96.33%
35	75.51%	75.43%	83.68%
36	91.37%	91.37%	93.11%
37	87.28%	87.44%	73.99%
38	95.75%	95.74%	94.87%
39	97.98%	98.07%	89.10%
40	93.54%	93.52%	95.81%
41	92.14%	92.23%	90.24%
42	95.05%	95.05%	94.76%
43	96.85%	96.63%	88.79%
44	91.14%	90.15%	94.49%
45	80.90%	80.13%	72.11%
Average	**92.12%**	**90.00%**	**89.31%**

As seen from the Table 14, k-means clustering approach provides the best overall performance with an average accuracy of 92.12% for all 45 images. The minimum accuracy was 70.16% for Image 20 and maximum accuracy was 99.01% for Image 4. For some of the images fuzzy c-means based clustering and GMM based clustering approaches provided better results. To identify the pattern associated with these improvements, the analysis was carried out for

psoriasis lesion sub-groups. The Table 15 below provides the performance comparison for the

three approaches within each psoriasis lesion sub-group.

Table 15: Psoriasis Lesion Sub-Group Performance Comparison							
Fuzzy C-Means Clustering Algorithm				Gaussian Mixture Model (GMM) Clustering Algorithm			
Red Sub-Group	k-Means	Fuzzy C-Means	Difference (%)	Red Sub-Group	k-Means	GMM	Difference (%)
3	97.83%	97.83%	0.00%	3	97.83%	97.91%	0.08%
4	99.01%	99.01%	0.00%	4	99.01%	96.47%	-2.54%
6	98.24%	98.21%	-0.03%	6	98.24%	97.69%	-0.55%
8	97.62%	97.62%	0.00%	8	97.62%	82.03%	-15.59%
13	98.46%	98.49%	0.03%	13	98.46%	97.03%	-1.43%
20	70.16%	32.20%	-37.96%	20	70.16%	98.07%	27.91%
22	95.84%	40.21%	-55.63%	22	95.84%	95.00%	-0.84%
23	98.52%	96.94%	-1.58%	23	98.52%	96.60%	-1.92%
25	94.91%	94.87%	-0.04%	25	94.91%	95.99%	1.08%
38	95.75%	95.74%	-0.01%	38	95.75%	94.87%	-0.88%
39	97.98%	98.07%	0.09%	39	97.98%	89.10%	-8.88%
43	96.85%	96.63%	-0.22%	43	96.85%	88.79%	-8.06%
Average	**95.10%**	**87.15%**	**-7.95%**	**Average**	**95.10%**	**94.13%**	**-0.97%**
Scaling Sub-Group	k-Means	Fuzzy C-Means	Difference (%)	Scaling Sub-Group	k-Means	Fuzzy C-Means	Difference (%)
1	90.42%	90.48%	0.06%	1	90.42%	61.91%	-28.51%
8	87.44%	87.40%	-0.04%	8	87.44%	82.03%	-5.41%
12	98.00%	97.89%	-0.11%	12	98.00%	96.03%	-1.97%
14	95.09%	95.09%	0.00%	14	95.09%	94.00%	-1.09%
15	98.40%	98.40%	0.00%	15	98.40%	98.66%	0.26%
17	98.20%	98.20%	0.00%	17	98.20%	97.74%	-0.46%
18	96.06%	96.04%	-0.02%	18	96.06%	90.37%	-5.69%
19	97.72%	97.63%	-0.09%	19	97.72%	94.91%	-2.81%
26	95.27%	95.25%	-0.02%	26	95.27%	88.06%	-7.21%
34	88.83%	88.98%	0.15%	34	88.83%	96.33%	7.50%
35	75.51%	75.43%	-0.08%	35	75.51%	83.68%	8.17%
41	92.14%	92.23%	0.09%	41	92.14%	90.24%	-1.90%
42	95.05%	95.05%	0.00%	42	95.05%	94.76%	-0.29%
44	91.14%	90.15%	-0.99%	44	91.14%	94.49%	3.35%
45	80.90%	80.13%	-0.77%	45	80.90%	72.11%	-8.79%

Average	92.01%	91.89%	-0.12%	Average	92.01%	89.02%	-2.99%
Red+Scaling Sub-Group	k-Means	Fuzzy C-Means	Difference (%)	Red+Scaling Sub-Group	k-Means	GMM	Difference (%)
2	92.16%	91.88%	-0.28%	2	92.16%	93.90%	1.74%
5	81.09%	80.87%	-0.22%	5	81.09%	91.96%	10.87%
9	74.47%	73.80%	-0.67%	9	74.47%	21.79%	-52.68%
10	72.65%	76.28%	3.63%	10	72.65%	86.67%	14.02%
11	98.16%	98.10%	-0.06%	11	98.16%	85.69%	-12.47%
16	98.74%	98.60%	-0.14%	16	98.74%	86.00%	-12.74%
21	98.39%	98.09%	-0.30%	21	98.39%	98.35%	-0.04%
24	90.85%	90.67%	-0.18%	24	90.85%	88.87%	-1.98%
27	95.07%	93.16%	-1.91%	27	95.07%	76.15%	-18.92%
28	87.17%	87.17%	0.00%	28	87.17%	87.83%	0.66%
29	88.95%	89.29%	0.34%	29	88.95%	93.30%	4.35%
30	92.55%	92.64%	0.09%	30	92.55%	92.48%	-0.07%
31	91.35%	91.99%	0.64%	31	91.35%	82.18%	-9.17%
32	93.17%	93.43%	0.26%	32	93.17%	92.72%	-0.45%
33	97.13%	97.43%	0.30%	33	97.13%	97.86%	0.73%
36	91.37%	91.37%	0.00%	36	91.37%	93.11%	1.74%
37	87.28%	87.44%	0.16%	37	87.28%	73.99%	-13.29%
40	93.54%	93.52%	-0.02%	40	93.54%	95.81%	2.27%
Average	90.23%	90.32%	0.09%	Average	90.23%	85.48%	-4.75%

In Table 15, the percentage difference between the k-means clustering and fuzzy c-means clustering approach was evaluated. Additionally, percentage difference between the k-means clustering and gaussian mixture model-based clustering algorithm was evaluated. Average accuracy and average percentage difference within each psoriasis lesion sub-group was evaluated.
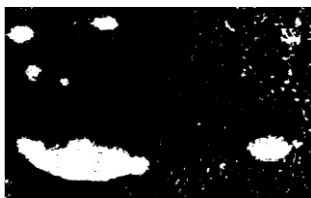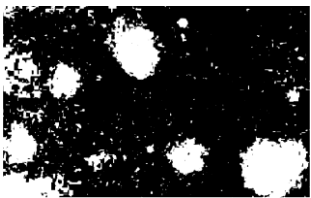
	Image 20	Image 5	Image 10
Original Image			
Ground Truth Image			
k-Means Clustering			
Segmentation Accuracy	70.16%	81.09%	72.65%
Fuzzy C-Means Clustering			
Segmentation Accuracy	32.20%	80.87%	76.28%
GMM Clustering			
Segmentation Accuracy	98.07%	91.96%	86.67%

Figure 21 – Visual Analysis of Segmentation Accuracy

As seen from the Table 15, fuzzy c-means clustering approach provides better results compared to k-means clustering based approach for the Red+Scaling sub-group but the increase in accuracy is 0.09% which is minimal. On the other hand, GMM based model offered significantly higher results compared to both k-means based clustering approach and fuzzy c-means based approach for some of the images. In fact, for three images the performance improvement was more than 10% which have been highlighted with orange color in the Table 15. These three images were further analyzed as shown in the Figure 21.

As seen from the Figure 21, for Image 20 the segmentation accuracy of GMM based clustering is far superior to both the approaches which fail to separate the individual lesions. One possible reason for this is due to the convex curve in the center of the image which indicates that the image is possible taken from the arms or legs of the individual and the k-means clustering, and fuzzy c-means based approach both fail because of it. Another thing to note is that GMM based clustering has a lot of noise besides the lesions which requires additional morphological operations. These additional morphological operations were intentionally not applied to ensure similar operations for all three algorithms for the sake of comparison.
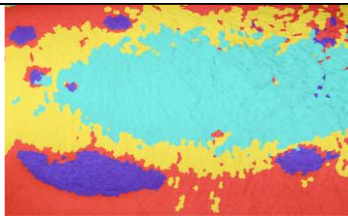
Superpixels based Segmentation Approach
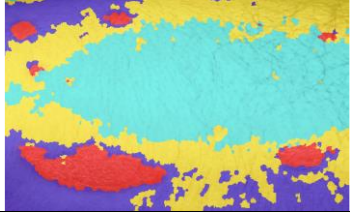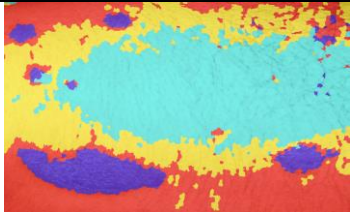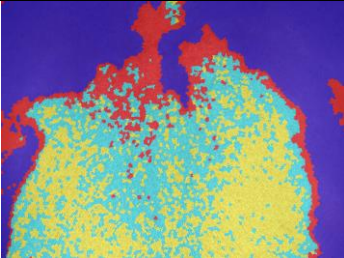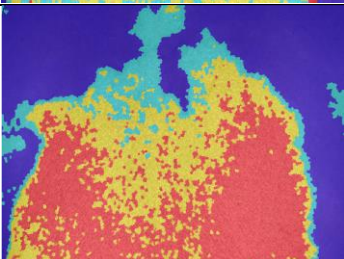
An approach for determining superpixels was proposed in [54], [59]. Superpixels were used to reduce the number of input pixels to the clustering algorithms by merging pixels based on similarity of color characteristics into larger superpixels. This reduces the effect of noisy

variations in color and thus improves the clustering ability of the algorithms. Following steps were adapted for the algorithm design:

- RGB to La*b* color-space conversion

- Determine number of superpixels as a ratio of number of pixels and a scalar 200. This means a reduction in number of pixels by a factor of 200.

- Determine a*b*, CbCr and IQ color-components

- k-Means was applied on all color components determined in the previous step.

- Number of K-means segments was taken as 4 (over-segmentation)

- Determine Mean and Standard Deviation of all 4 clusters for each of the color spaces.

- Merge clusters together with lower mean difference.

- Combine segmentation results of all three color-spaces.

- Apply Majority Voting to determine result.

The Figure 22 shows some of the results.

Image 20		
Color component	**k-means clustering**	**Merged clusters and generated mask**
a*b* components		

CbCr components		
IQ components		
	Ground Truth	Result – Majority Voting
		 Accuracy ~ 97.46%

Image 19

Color component	k-means clustering	Merged clusters and generated mask
a*b* components		
CbCr components		

IQ components		
	Ground Truth	Result – Majority Voting
		 Accuracy ~ 96.83%
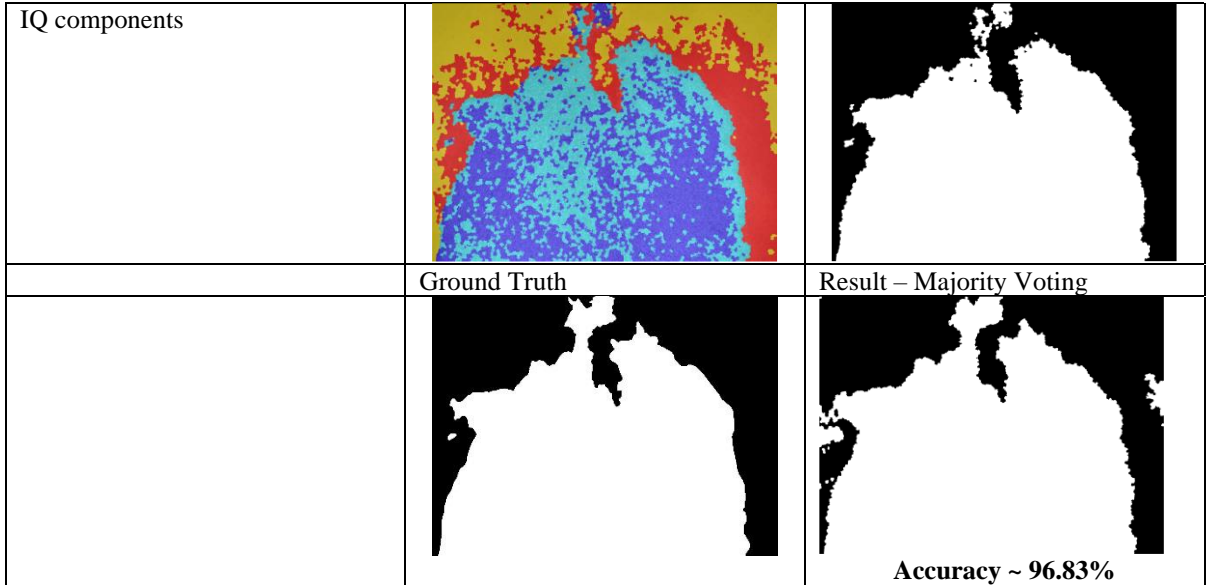
Figure 22: Proposed Superpixels based Clustering Algorithm

The Table 16 summarizes the results across all the images in the dataset.

Table 16: Performance Comparison of superpixels based approach and earlier k-means based approach			
Image Number	**Accuracy (Superpixels)**	**k-Means Approach (Figure 13(a))**	**Performance Decreased**
Image 1	93.59%	90.42%	
Image 2	92.31%	92.16%	
Image 3	97.93%	97.83%	
Image 4	98.72%	99.01%	-0.29%
Image 5	96.17%	81.09%	
Image 6	98.19%	98.24%	-0.05%
Image 7	96.81%	97.62%	-0.81%
Image 8	89.74%	87.44%	
Image 9	77.68%	74.47%	
Image 10	94.76%	72.65%	
Image 11	98.09%	98.16%	-0.07%
Image 12	95.40%	98.00%	-2.6%
Image 13	97.65%	98.46%	-0.81%
Image 14	92.62%	95.09%	-2.47%
Image 15	96.92%	98.40%	-1.48
Image 16	98.55%	98.74%	-0.19
Image 17	96.75%	98.20%	-1.45
Image 18	93.67%	96.06%	-2.39

Image 19	96.83%	97.72%	-0.89
Image 20	97.46%	70.16%	
Image 28	91.81%	87.17%	
Image 29	93.03%	88.95%	
Image 34	97.92%	88.83%	
Image 35	88.69%	75.51%	
Image 37	94.46%	87.28%	
Image 45	84.29%	80.90%	
Average	**94.23%**	**90.33%**	

As seen from Table 16, superpixel based approach provides superior performance to k-means based approach and it relies on more color features thus extracts more information for the same lesions.

Chapter Conclusion

The focus of the research work presented in this chapter was to accurately segment psoriasis lesions from normal skin and to determine the color features that can be used with a linear SVM classifier to determine region properties. The k-means clustering approach with morphological operations provided accurate segmentation of the psoriasis lesions with the best overall average accuracy. The FCM algorithm provided the best average accuracy for the Red+Scaling sub-group but the increase in accuracy was not significantly high. The GMM algorithm provided lower overall average accuracy but for some of the images it outperformed both the other algorithms. Considering the performance gain achieved from the GMM based proposed algorithm, it can be used in a semi-automatic environment where the end user can be asked to confirm visually which output is better between the k-means and the GMM in terms of proper lesion segmentation. Alternatively, a weighted average approach may be used to

consider output from all three segmentation approaches with the final decision based on majority voting. The maximum accuracy achieved in such a case was found to be 94%.

The results also demonstrated that a small set of statistically significant features can accurately distinguish between psoriasis and normal skin for red and red + scaling sub-groups. However, for the scaling sub-group the classifier has higher predictive inaccuracy due to the reduced separability between lesion and normal skin for the scaling group. This requires a larger dataset for classifier training to avoid overfitting the model that may happen if a higher order kernel is used for the classifier.

A limitation of the current study was the use of small image database as there are no known large image databases publicly available for psoriasis research. In the next chapter, a review of deep learning-based approaches to biomedical image segmentation will be presented.

DEEP LEARNING APPROACHES TO BIOMEDICAL IMAGE SEGMENTATION

In this chapter, a review of automatic segmentation of images by means of deep learning approaches in medical imaging is presented. Current developments in deep learning, are proving instrumental in identification, and quantification of patterns in the medical images. The pivotal point of these advancements is the essential capability of the deep learning approaches to obtain hierarchical feature representations directly from the images, which in turn is removing the need for hand crafted features. Deep learning is expeditiously turning into the state of the art for medical image processing and has resulted in performance improvements in diverse clinical applications. In this review, the basics of deep learning methods are discussed along with an overview of successful implementations involving image segmentation for different medical applications. Finally, some research issues are highlighted. This chapters provides the essential background and literature review for the deep learning approaches to psoriasis image segmentation techniques presented in subsequent chapters.

Introduction

There are two components of medical imaging: 1) image formation and reconstruction and 2) image processing and analysis [60]. The image formation involves the set of processes through which two dimensional (2D) images of three dimensional (3D) objects are formed while reconstruction relies on a set of iterative algorithms to form 2D and 3D images typically from the projection data of an object. Image processing, on the other hand, involves the use of algorithms to enhance image properties like noise removal while image analysis extracts

quantitative information or a set of features from the image for object identification or classification.

The ease of image acquisition due to advancing technology has led to generation of vast amounts of high-resolution images at very low costs. This has led to a significant improvement in the development of biomedical image processing algorithms. This has in turn enabled development of automated image analysis or evaluation algorithms to extract useful information. The basic step for such automated analysis is segmentation which sub-divides the image based on visually distinct regions which have a semantic meaning for the given problem. Each of these region typically have uniform characteristics either in terms of their gray level, texture or color [61]. Clear segmentation and distinguishable regions are essential for further analysis that may involve determination of homogeneity levels of texture or layer thickness [62]. At times the image may contain multiple objects of the same class. And the segmentation process may segregate regions containing the objects of the same class while ignoring other classes in which case it is known as instance segmentation as opposed to semantic segmentation in which objects of the same class are not segregated but different classes are segregated.

All image segmentation techniques can be grouped into three categories: 1) Manual segmentation (MS), 2) Semi-automatic segmentation and 3) fully automatic segmentation techniques [63]. MS techniques require subject experts to first determine the region of interest (ROI) and then draw precise boundaries surrounding the ROI to correctly annotate each of the image pixels. MS is necessary as it provides the ground truth labeled images for the further

development of semi-automatic and fully automatic segmentation techniques. MS is time intensive, only feasible for smaller image datasets. In case of high-resolution images, the high resolution may result in images no longer having a crisp boundary (weak contrast) resultantly slight variations in selection of pixels of ROI boundary can result in a large error. Another issue with the manual segmentation is that it is subjective as the approach is dependent on the expert's knowledge and experience and as a result often encounters significant inter and intra expert variability [64]. Semi-automatic segmentation techniques involve small level of user interactions with automated algorithms to produce accurate segmentation results. The user interaction may involve selection of approximate initial ROI which is used subsequently to segment the entire image. It may involve manual checking and editing of region boundaries to reduce segmentation error. Examples of semi-automatic segmentation techniques include: 1) seeded region growing (SRG) algorithm that iteratively merges neighborhood pixels with similar intensity based on user provided initial seed point [65], 2) level-set based active contour model which starts with initial boundary shapes represented by contours and iteratively alters them through shrinking or expansion operation based on the implicit level of a function and has the advantage that it does not require prior shape knowledge and the initial locations of the ROI [66] and 3) localized region-based active contour techniques that utilizes region parameters to describe foreground and background of the image using small local regions and has the advantage of handling heterogeneous textures [67].

The fully automatic segmentation techniques do not require any user interaction. However, most such techniques are based on supervised learning approaches that require training data,

e.g., shape models, atlas-based segmentation approaches, random forest and deep neural networks. Both the training data and in case of unsupervised learning approaches the validation data requires labeled images which are obtained through manual segmentation thus imposing similar constraints as mentioned earlier. Additional challenges with the automated segmentation of medical images is the large variations in shape, size, texture and in certain cases color of ROI between patients and poor contrast between regions [68]. Noise or lack of consistency in source data acquisition may also result in wide variations in the source image data which is often the case in real applications. Due to this reason, most existing approaches based on clustering techniques, watershed algorithms and machine learning (ML) based approaches have the basic problem of lack of global applicability that limits their use to narrow limited number of applications for which they were originally intended. Furthermore, human feature engineering which are often utilized with machine learning approaches based on support vector machines (SVM) or neural networks (NN) are time consuming and fail to handle natural data in their raw form and typically do not adapt to new information. Deep learning approaches on the other hand are capable of processing natural data in their raw form thus eliminating the need for hand-crafted features [61]. These approaches have been effectively used for semantic segmentation on images of nature and have also found applications in biomedical image segmentation [69]. The increased usage of deep learning approaches has been facilitated due to faster central processing units (CPUs) and graphic processing units (GPUs) that have largely reduced the training and execution time, access to large sets of data and advances in learning algorithms [70]. The remaining parts of this chapter is structured as

follows. In next section, an overview of machine learning approach to image segmentation, deep learning approach to image segmentation, deep learning architectures, typical approaches for implementing deep learning architectures and performance metrics used for image segmentation is provided. Then, recent studies using deep learning models for different biomedical image segmentation applications are presented which is followed by a discussion of the studies presented in previous section. And finally, the conclusion is provided that summarizes research trends, challenges associated with deep learning-based image segmentation.

Machine Learning

The typical application of machine learning based image segmentation approach is to classify the ROI e.g., diseased region or healthy region. The steps involved for designing such an application begins with the pre-processing stage which may involve the use of a filter to remove any noise or for the purpose of contrast enhancement. Following the pre-processing stage, the image is segmented using a segmentation technique like thresholding, clustering based approach and edge-based segmentation. After the segmentation, features are extracted based on color information, texture, contrast and size from the ROI. Then dominant features are determined using feature selection techniques like principal component analysis (PCA) or statistical analysis. Subsequently the selected features are used as an input to the ML classifier such as SVM or NN. The ML classifier uses the input feature vector along with target class labels to determine the optimal boundary that separates each class. Once the ML classifier has been trained, it can be used to classify new unknown data to determine its class. Typical

challenges include determination of appropriate pre-processing requirements based on raw image properties, determination of the appropriate features and length of the feature vector and the type of classifier, among others. The methodology for psoriasis image segmentation is an example of machine learning application to biomedical image processing.

Deep Learning-based Classifier (DLC)

DLC can process raw image directly which means, there should be no need for pre-processing, segmentation, and feature extraction. However, most deep learning approaches require image resizing due to limit on input values. While some techniques do require intensity normalization and contrast enhancement which may be avoided if data augmentation techniques discussed later are employed during training. Resultantly, DLC has higher classification accuracy as it can avoid errors associated with erroneous feature vector or imprecise segmentation [71]. The comparison of ML and DLC approaches is shown in Figure 23 below. DLC based approaches have altered the focus of research from traditional image processing techniques for feature engineering to network architecture design for obtaining optimal results. DLC networks typically have multiple hidden layers which resultantly means more mathematical operations are performed compared to ML based approaches and thus the models are more computationally intensive.
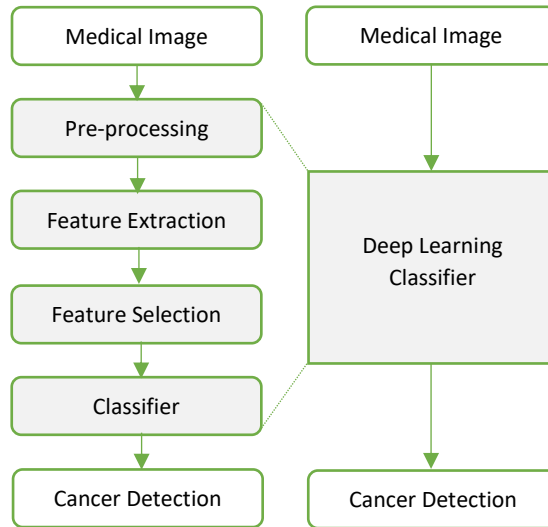
Figure 23. Change in classifier approach using typical machine learning algorithm and deep learning. Figure adapted from Ref. [71].

As seen from Figure 23, the machine learning classifier takes in the feature vector as input and the output is the object class while the deep learning classifier takes in the image and the output is the object class. It may be noted that theoretically deep learning can be said to be an enhancement of conventional artificial neural networks (ANN) as it consists of more layers than ANN [70]. It is considered as a type of representational learning as each layer transforms the input data from the previous layer into a new representation at a higher and somewhat greater abstraction level [61]. This allows the model to learn both local and inter-relationships of the whole data in a hierarchical structure. The transformation of data into representation in each layer of a deep learning model is a result of non-linear function. Usually, features extracted from the first layer of representation for a given image will identify the presence or absence of edges in specific alignment and its location in the image. The second layer detects

patterns by recognizing positioning of edges and ignores minor differences in these positions while the third layer associates these patterns into bigger combinations corresponding to fragments of similar objects while enabling succeeding layers to detect objects through these combinations. This hierarchical feature representation learning directly from the data has led to unprecedented success of deep learning in a range of artificial intelligence applications [70].

Deep Learning Architecture – Convolutional Neural Network (CNN)

Among the many deep learning architectures, the CNN is the most widely used as it is very similar to conventional NN. As opposed to a typical NN, shown in Figure 24(a), CNN takes in an image as input and have three-dimensional arrangement of neurons that connects with a small region of the preceding layer instead of the entire layer as shown in Figure 24(b) below. The CNN comprises of layers which include convolutional layer, non-linear activation layer such as rectified linear unit layer (ReLU), pooling layer or fully connected layer. The convolutional layer applies convolution operation between pixels of the input image and a filter to obtain volumes of feature maps containing features extracted by the filter. ReLU is a non-linear activation layer that applies the function $f(x) = \max(0, x)$ to the input values to increase non-linearity and improve the training speed. Pooling layer down-samples the input values to reduce the spatial dimensionality of the image to improve computational cost and prevent overfitting and are translation invariant as computations are based on neighboring pixels [72]. Fully connected layer is typically the last layer of CNN and is like the hidden layers of traditional NN in the sense that all neurons in this layer are linked to the neurons in the preceding layer. As mentioned earlier that CNN is typically used for classification problems.

To use CNN for semantic segmentation, the input image is sub-divided into small patches of equal size. The CNN classifies the center pixel of the patch. Then the patch is glided forward to classify the next center pixel. However, such an approach is inefficient as the overlapping features of the sliding patches are not re-used resulting in loss of spatial information of the image as the features move into the final fully connected network layers. To overcome this problem the use of fully convolutional network (FCN) was proposed in which the final fully connected layers of the CNN were changed to transposed convolutional layers, as shown in Figure 24(c), which applies up-sampling on the low resolution feature maps to recover the original spatial dimensions while simultaneously performing semantic segmentation [68].

Typically, deep neural networks are trained using backpropagation algorithm in combination with an optimization algorithm like gradient descent. The process involves determination of the gradient of a loss function which is used by the optimization algorithm to update the network weights to minimalize the loss function value.

Other Architectures

Restricted Boltzmann Machines (RBMs)

Restricted Boltzmann Machines (RBMs) are neural networks designed on the principles of energy-based models (EBMs). EBMs encode reliance between variables by assigning a scalar energy to individual arrangement of the variables. Inference or prediction is made using values of observed variables to determine the values of the residual variables in a manner to minimize the energy.

Input Layer

Output Layer
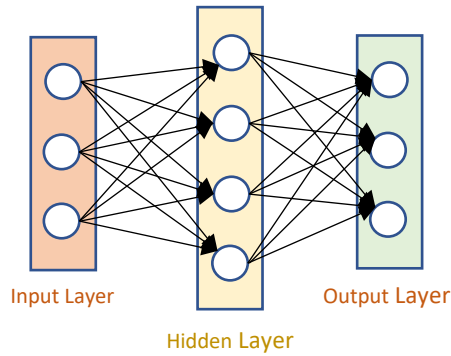
Hidden Layer

(a) Traditional Neural Network
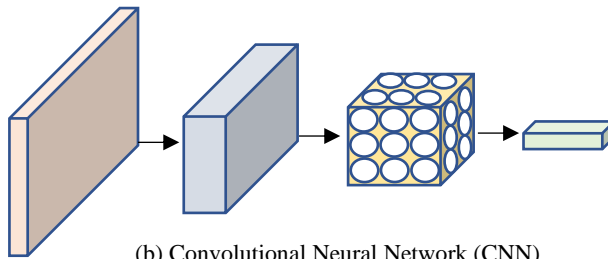
(b) Convolutional Neural Network (CNN)
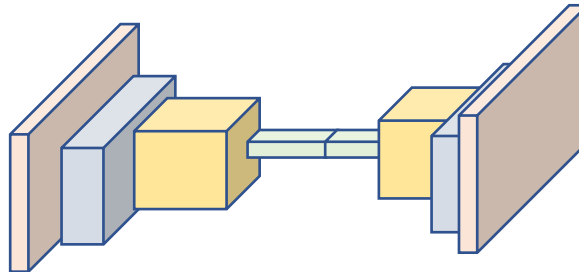
(c) Fully Convolutional Network (FCN)

Figure 24(a). A 2-layer Neural Network (one hidden layer of 4 neurons and one output layer with 3 neurons), and three inputs, (b). Convolutional Neural Network (CNN) and (c) Fully Convolutional Network (FCN)

While learning is achieved through determining an energy function that yield minimal energy for the right values of the residual variables and larger energies for erroneous values. The loss function minimized during learning, in turn provides a measure of the superiority of the available energy functions. The RBM consists of one input layer (I1, …, I4), one hidden layer

(h1, h2), a bias vector (b1, b2) and a weight vector *(w)* but does not have an output layer. A simple architecture of RBM is shown in Figure 25.

Based on the architecture shown in Figure 25, the energy function for the RBM can be defined as,

$$E(I, h) = -\sum_i a_i I_i - \sum_j b_j h_j - \sum_{i,j} I_i h_j w_{i,j} \tag{12}$$

During the training of RBMs, network parameters are determined for given inputs that minimizes the energy function provided in Eq. 12. RBMs are probabilistic models and the values of neurons in input and hidden layers implies the state at a specific point in time and these values indicate if the corresponding neuron is either active (state 1) or inactive (state 0).
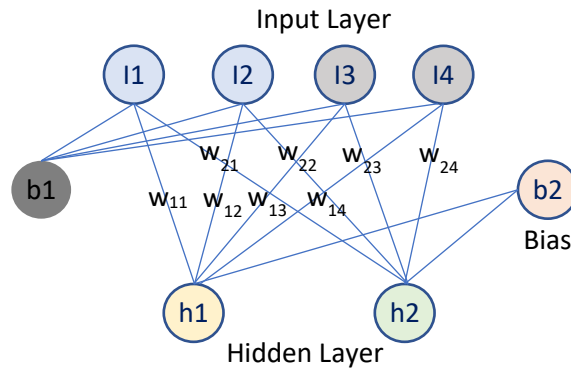
Figure 25. Restricted Boltzmann Machines (RBMs)

A type of RBM achieved through stacking is known as Deep Belief Network (DBN) in which each layer communicates with preceding and subsequent layers. The top two layers contain undirected connections while lower layers have directed connections. As opposed to another

type of RBM network known as Deep Boltzmann Machine (DBM) which only have undirected connections. DBMs are considered to handle uncertainty better in the presence of noisy inputs.

Autoencoder based Deep Learning Architectures

An autoencoder neural network is an unsupervised learning algorithm that compressing the input into a latent-space representation by applying backpropagation algorithm with target values same as the inputs. It comprises of two parts: 1) Encoder portion of the network that condenses the input into a latent-space representation expressed by function $h = f(x)$, and 2) Decoder part of the network that reconstructs the input from the latent space representation. The compression is achieved by constraining the hidden layer to have lower dimensions than the input layer. Such a network is known as undercomplete. Lower dimensionality of the hidden layer results in network learning the most prominent features in the training data. Alternatively, a sparsity constraint can also be applied to achieve similar results by keeping neurons in the hidden layer inactive most of the time. In autoencoder based deep learning approaches, the input comprises of the image which is down sampled to obtain latent representation of lower dimensions enabling the autoencoder to be trained and learn on condensed form of the images. The autoencoder architecture is shown in Figure 26.

One of the challenges in the autoencoders is that nodes in the hidden layer are greater than the number of input values. This has the risk of network learning null or identity function in which the output equals the input. To resolve this issue denoising autoencoders are used in which the data is intentionally distorted by randomly assigning around 30 to 50% of the input values as zero. The actual values reduced to zero depend on the data size and number of nodes

present in the network. When determining the loss function, the output is compared with the original input thus the risk of null function learning is eliminated.

The applications of autoencoders are relatively limited due to discontinuities in the latent space representations that do not allow its application as a generative model. To resolve this issue, variational autoencoders were introduced. In variational autoencoder, the output of the encoder is not a single encoded vector rather it outputs two encoded vectors; one is a vector of means and the other is a vector of standard deviations. These vectors act as the parameters of a random variable from which the output encoded vector is sampled. This allows the decoder to correctly decode the encoded values even in the presence of small variations of the same input during training. The stochastic nature of the autoencoder ensures that the latent space representation is by design continuous thus allowing for random interpolation and sampling.

Sparse Coding based Deep Learning Architectures

Sparse coding is a class of unsupervised learning that determines an overcomplete set of basis vector to represent the input data. Overcomplete means that the dimension of the latent representation is higher than the input. The aim is to determine a linear combination of these basis vectors corresponding to a given input. The overcomplete network means that additional constraint of sparsity needs to be applied to address any degeneracy. The advantage of sparse coding is that it can identify correlations between similar descriptors and capture salient properties of images [72].

Generative Adversarial Networks (GANs)

In GANs, the idea is to have a generator implemented through neural network such that it models a transform function which takes in a random variable as input and follows the targeted distribution when trained. While another network is trained as a discriminator simultaneously to distinguish between generated data and true data. The two networks operate as adversaries that is the first network tries to maximize the final classification error between generated data and true data while the other tries to minimize the same error. Resultantly, both networks improve after each iteration of the training process.
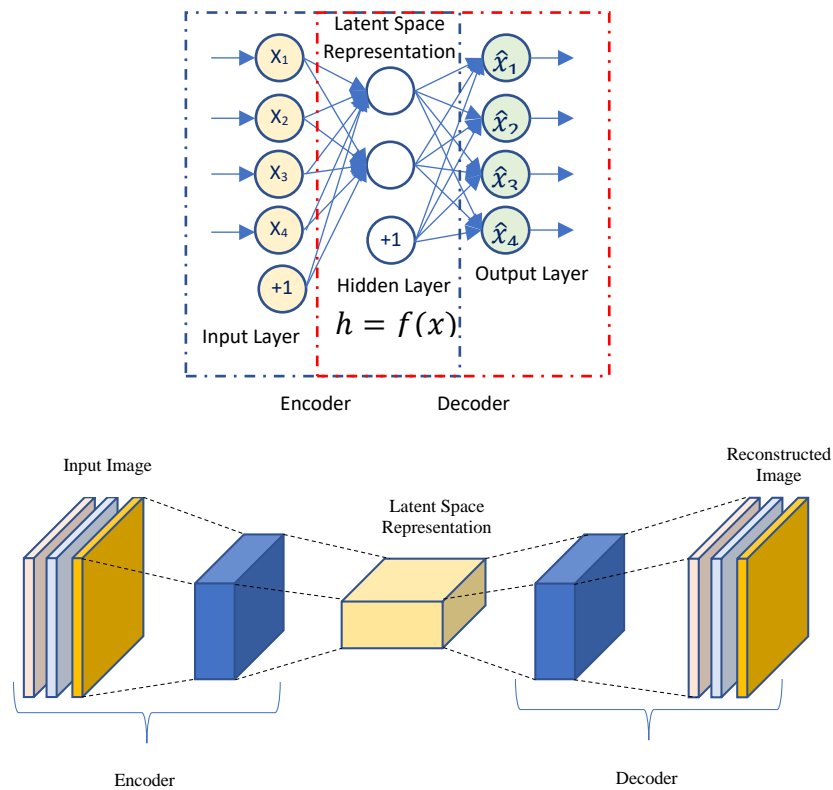
Figure 26. Autoencoder architecture with vector and image inputs

Recurrent Neural Networks (RNNs)

RNNs are specifically designed to operate with series type inputs when the size of the input cannot be pre-determined. The series input is different from many inputs as it impacts its neighboring values and therefore this relationship must be recognized by the network. RNNs are such networks whose present output is based on both the current input and learning based on past values. The prior input information is part of the network and stored in a hidden state vector. This means the same input can result in different output depending upon the previous inputs in the series. The network becomes recurrent when repeatedly transformed with different input series values resulting in generation of different fixed size output vectors. The hidden state is updated with each input. The depth can be added to RNNs by adding more hidden state layers or by adding non-linear hidden layers between the input and the hidden state layer or by adding more layers between hidden state layer and the output layer or use a combination of all three.

Typical Approaches for Implementing Deep Learning Architectures

There are multiple approaches through which deep learning techniques have been implemented for image segmentation. In the first approach the neural network is trained from scratch which usually requires availability of large, labelled dataset and is time intensive to build and train the network. In the second approach, one of the existing pre-trained CNNs like AlexNet which was trained to classify 1.2 million high-resolution images, available through the ImageNet Large Scale Visual Recognition Challenge 2010 for 1000 different classes, can be used [73]. The typical process in such an approach is to remove the last few layers of the

network and replace them with the new task specific layers. The low-level features learned from millions of images in the beginning layers are merged with the task specific features extracted in the final layers to implement the network for classification of new images. This provides the advantage of reduced time for implementation as only a small number of weights need to be found, respectively. Transfer learning is typically employed with networks trained on data from ImageNet and is better than random initialization of weights [74].

The third approach involves the use of pre-trained CNNs for extracting features from the data and subsequently employing those features as inputs for training a traditional classifier like support vector machine for classification. The advantage of this approach is that the features can be automatically extracted for large number of categorical data thus eliminating the time-consuming need of human engineered feature extraction.

Some of the well-known convolutional neural networks include U-Net that was developed for biomedical image segmentation [75] and V-Net that was developed for volumetric medical image segmentation [76]. U-Net is a type of FCN with a contraction path and expansion path. The contraction path consists of consecutive convolutional layers and max pooling layer. It is used to extract features while limiting the size of feature maps. Expansion path performs up-conversion and has convolutional layers to recover the size of segmentation map with loss of localization information. Skip connections are used to share localization information from the contraction layer to the expansion layer. These are parallel connections allowing signals to propagate directly from one block of the network to another without adding any computational complexity. Finally, the convolutional layer before the output maps the feature vector to the

required number of target classes in the final segmentation output. V-Net is similar to U-Net and consists of two parts: 1) compression part and 2) decompression part. The compression part consists of multiple stages with each stage consisting of 1 to 3 convolutional layers. At each stage residual function is learned with convolution operations performed using volumetric data based on voxels. Compression path reduces the resolution by half through convolution like pooling layer. However, pooling layer is not used to reduce memory utilization. Parametric Rectified Linear Unit (PReLU), also known as Leaky ReLU, is a generalization of ReLU and is used as non-linearity activation function. The decompression part of the network expands the spatial support of the feature maps to produce sufficient information for volumetric segmentation. Deconvolution is used to increase the size of the inputs and residual function is learned like the compression part of the network. The convolutional layer before the output produces two feature map outputs with the same size as input volume. The two feature maps contain the predicted foreground and background region information. Skip connections are used in a similar manner to U-Net to forward localization information from the contraction part of the network to the expansion part of the network.

Performance Metrics

The efficacy of the image segmentation system is evaluated by using standard and well-known metrics which enables comparison of the system with existing techniques in literature. Selection of appropriate evaluation metric depends on many factors which primarily depend on the functionality of the system. These metrics may measure computational complexity, processing time, memory utilization and accuracy among others [74]. Various performance

metrics are provided below that can be used to assess the segmentation efficacy of the deep learning models in terms of accuracy. TP, FP, FN, and TN are defined in the Table 17 below.

Table 17: Definition of the Abbreviations		
Category	Actual Disease	Actual No Disease
Predicted Disease	True Positive (TP)	False Positive (FP)
Predicted No Disease	False Negative (FN)	True Negative (TN)

Accuracy

Accuracy is defined as,

$$Accuracy = \frac{Correctly\ Predicted\ Pixels}{Total\ number\ of\ Image\ Pixels} \tag{13}$$

$$= \frac{TP + TN}{TP + FP + FN + TN}$$

which represents the percent of image pixels that are classified correctly. It is also known as overall pixel accuracy. It is the most basic performance metric but has the limitation to misrepresent image segmentation performance in case of class imbalance. Class imbalance occurs when one segmentation class dominates the other. In such a case, higher accuracy for the dominating class will overshadow the lower accuracy associated with the other class thus providing biased results. Therefore, the accuracy measure is recommended to be used for evaluating segmentation performance with images when there is no class imbalance.

An alternative to above definition of accuracy is per-class accuracy that determines the percent of correctly labelled pixels for each class and then determines their average [77]. This measure is useful for images where class imbalance is present. In case of class imbalance both the average and the per-class accuracy must be considered. The limitation of the average per-class

96

accuracy is the reduction in confidence measurement of the individual classes. Furthermore, fewer instances of a given class may result in large variance thereby impacting the reliability of the results.

Precision

Precision is defined as,

$$Precision = \frac{Correctly\ Predicted\ Disease\ Pixels}{Total\ number\ of\ Predicted\ Disease\ Pixels} \tag{14}$$

$$= \frac{TP}{TP + FP}$$

and represents the proportion of the disease pixels in the automatic segmentation results that match with the ground truth disease pixels. Precision is a useful measure of segmentation performance as it is sensitive to over-segmentation. The over-segmentation results in low precision scores.

Recall

Recall is defined as,

$$Recall = \frac{Correctly\ Predicted\ Disease\ Pixels}{Total\ number\ of\ Actual\ Disease\ Pixels}$$

$$= \frac{TP}{TP + FN} \tag{15}$$

and represents the proportion of disease pixels in the ground truth that were correctly identified through automatic segmentation. It is sensitive to under-segmentation as those results in low recall scores.

F1$_{measure}$

Precision and recall can be jointly used, as high values for both measures for a given segmentation result means that the predicted segmented regions match with the ground truth both in terms of location and the level of detail. $F1_{measure}$ also known as Boundary F1 (BF) determines the harmonic mean of precision and recall and is useful for contour or boundary matching between predicted segmentation and ground truth segmentation. It is defined as,

$$F1_{measure} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{16}$$

It is also known as the Dice Similarity Coefficient (DSC) whose alternative definition is provided below.

DICE Similarity Coefficient (DSC)

DSC is defined as,

$$\begin{aligned} DICE &= \frac{2|S_{Ground\ Truth} \cap S_{Automated}|}{|S_{Ground\ Truth}| + |S_{Automated}|} \\ &= \frac{2 \times TP}{2 \times TP + FP + FN} \end{aligned} \tag{17}$$

where S stands for segmentation. As seen from the equation, DSC considers both the false alarm and the missed value in each class thus is superior to overall pixel accuracy. DICE is

also considered to be superior as it not only evaluates the number of pixels correctly labeled but also determines the accuracy of the segmentation boundaries [77]. Additionally, DICE is often used to measure repeatability of system performance through cross-validation.

Jaccard Similarity Index (JSI)

Jaccard Similarity Index (JSI) is also known as Intersection-Over-Union (IoU) is defined as,

$$JSI = \frac{S_{Ground\ Truth} \cap S_{Automated}}{S_{Ground\ Truth} \cup S_{Automated}} = \frac{TP}{TP + FP + FN} \tag{18}$$

where S stands for segmentation. Alternatively, it's the ratio of the area of the overlap between the predicted segmentation and the ground truth segmentation to area of union between the predicted segmentation and the ground truth segmentation. JSI is like DSC as they are monotonic in one another or positively correlated as can be seen from equations,

$$JSI = \frac{DSC}{2 - DSC} \tag{19}$$
$$DSC = \frac{2JSI}{1 + JSI}$$

The difference between JSI and DSC as seen from the above equation is that JSI will penalize instances of incorrect results more than the DSC. Thus, any one of these two metrics can be used for segmentation validation instead of using both measures.

Modified Hausdorff Distance (MHD)

Given two sets of pixels A and B, the MHD is defined as,

$$H(K,G) = \max\left\{\underset{a\,\epsilon\,A}{max}\left\{\underset{b\,\epsilon\,B}{min}\{D(a,b)\}\right\}, \underset{b\,\epsilon\,B}{max}\left\{\underset{a\,\epsilon\,A}{min}\{D(a,b)\}\right\}\right\}$$

(20)

where $D(a, b)$ denotes the Euclidean distance between pixels a and b.

Small value of MHD specifies greater adjacency of the two-point sets, thus suggesting superior segmentation performance [78]. Sometimes k[th] ranked distance is determined based on Hausdorf Distance (HD) which is defined as,

$$HD(G,S) = \max\{h_{95}(S,G), h_{95}(G,S)\},$$

(21)

where,

$$h_{95}(S,G) = {}^{95}K_{s\epsilon S}^{th}\underset{g\,\epsilon\,G}{min}\|g - s\|$$

and G represents the ground truth.

Absolute Volume Difference (AVD)

AVD is defined as,

$$AVD(G,S) = \frac{|V_s - V_g|}{V_g}\ 100\%$$

(22)

where V_s is the volume of segmentation results and V_g represents the volume of ground truth. A smaller value of $AVD\ (G, S)$ implies a greater segmentation accuracy [79]. AVD and MHD are sensitive to point positions and thus can evaluate the segmentation accuracy more effectively when exact boundary delineation is needed to be determined [80]. Another

advantage of using MHD and AVD is that these measures do not penalize low density segmentation results as strictly as compared to other measures [80]. Low density means that the image segmentation has more accurate boundary contours but has several tiny holes.

For a more detailed reading on segmentation metrics, following literature articles can be consulted [74], [77], [80]–[82].

Types of Biomedical Images:

There are various types of biomedical images that are dependent on the method of imaging. Some of the widely used biomedical imaging techniques are provided below. The list below is not exhaustive as with advancement in technology new imaging techniques are being introduced to achieve better and timely diagnosis.

(i)   Clinical Images

Clinical images are digital images of the patient's body and are often used to document injury, burn, or skin lesions. The automatic analysis of these images may be used to track the efficacy of treatment over time. These images are widely used for dermatological and cosmetic treatments to track before and after representation of the skin or anatomical structure. The most widely used application of clinical images is for the detection of skin cancer known as melanoma.

(ii)   X-ray Imaging

X-ray imaging is the most widely used imaging technique to detect fractures and bone dislocation. The generated image is two-dimensional. The National Institutes of Health (NIH)

has provided open access to 100,000 chest x-ray images with associated data and diagnoses for improving imaging analysis techniques [83]. Similarly, Massachusetts Institute of Technology (MIT) has published a dataset containing a collection of more than 350,000 chest x-rays for developing machine learning models to automatically detect 14 of the common illnesses like pneumonia or punctured lung etc. [84].

(iii)   Computed Tomography (CT)

CT refers to computerized imaging procedure in which x-rays are aimed at the patient in all 360 degrees to produce detailed cross-sectional images of the internal organs, bones, soft tissue and blood vessels in the body. The images are traditionally captured in the axial or transverse plane and perpendicular to the long axis of the body. However, these images which are also known as slices can be reformatted into multiple planes and can generate three-dimensional image. It is widely used to detect cancer by localizing the presence of tumors and its size and is one of the most widely tackled biomedical imaging problem. The National Institutes of Health (NIH) has provided open access to 32,000 CT images with associated data and diagnoses for improving lesion's recognition accuracy [85].

(iv)   Magnetic Resonance Imaging (MRI)

MRI is an imaging technique that is used to form images of physiological processes, organs and tissues within the body using strong magnetic fields. MRI is used to image the non-bony parts or soft tissues of the body. The main difference from CT scans is that it does use the ionizing radiation of x-rays. Knee and shoulder injuries can be seen with better resolution in MRI scans as compared to both x-ray and CT scans. In the brain, MRI scans can be used to

distinguish between grey and white matter which in turn helps with determining aneurysms and tumors. For biomedical imaging researchers, Open Access Series of Imaging Studies (OASIS) project has gathered neuro-imaging datasets which contain more than 2000 MRI sessions [86].

(v)   Ultrasound Imaging (US)

US imaging technique uses high-frequency sound waves to produce visual images of the internal organs, tissues, and blood flow. It is the most widely used technique to monitor fetus during pregnancy. It is mostly used for abdominal, vascular, and thyroid scans and typically not used for imaging bones or tissues that contain air, like lungs. The benefit of using US is that it is fast and radiation free.

(vi)   Optical Coherence Tomography (OCT)

OCT is a technique that employs low-coherence light to acquire micrometer-resolution, two- and three-dimensional images from within biological tissue. OCT is primarily used for diagnosing eye problems by providing cross-sectional view of retina enabling the physician to distinctly see each layer. This enables layer mapping and measurement of thickness which is useful for the diagnosis.

(vii)   Microscopic Images

Microscopic medical images are used to analyze the microscopic structure of the tissue. The tissue to be analyzed is usually obtained through biopsy and then sections of the tissue are dyed with staining components to reveal details at cellular level. Counterstains are used to provide color, visibility, and contrast for the images. These images are widely used for

detection of cancer. The features usually analyzed include shape and size of the cells and its nucleus and distribution of the cells in the tissue.

Data Augmentation

The performance of deep learning neural networks is dependent on the availability of sufficient data. The problem is in most cases the training samples are not available in the required numbers especially in medical imaging. In order to enhance the available datasets in the absence of real data, various data augmentation techniques are employed for creating additional training data from the existing available dataset. The techniques for data augmentation apply class preserving transformations on the image data and may include: 1) shifting image pixels in one direction either horizontally or vertically without any change in the overall image dimensions (image translation); 2) horizontal and vertical flip of image pixels by reversing the rows and columns of pixels (image flipping); 3) rotation of image by certain degree from 0 to 360 (image rotation); 4) varying the brightness level of the images to train the model to account for such variations in test images; 5) zoom in or out of the image randomly either by the addition of new boundary pixels or through interpolation. In most of these techniques, some of the existing pixels are discarded and some new pixels are added either through nearest neighbor fill, duplication of boundary pixels, averaging or interpolation.

The first four techniques discussed above are known as rigid data augmentation techniques that is the shape itself remains unchanged. In the fifth technique, the ratio of horizontal and vertical augmentation is kept the same. However, if it is different than the image will stretch more in one direction compared to the other (image stretching). Alternatively, if the image is

stretched in only one direction along the diagonal axis on both ends, the image will be sheared (image shearing). Another technique which is known as elastic deformation can also be used which produces an effect in the shape of the region of the interest which is equivalent to stretching under external force. The change in shape is like how solid objects deform under external stress that is recoverable after the stress is removed. Lastly, contrast enhancements can be applied to adjust intensity variation in the image as medical images may have been obtained from a variety of sources.

The primary purpose of these augmentation techniques is to improve the generalizability of the deep neural network while avoiding both the underfitting and overfitting of the features. These techniques are usually applied automatically during the training phase of the network. Furthermore, in most cases linear transformations are sufficient and a safer approach to use. While heavy augmentations may produce variations in features, which are not realistic. However, the final choice again will be largely dependent on the type of medical image and properties of the region of interest. Lastly, data augmentation techniques can be used to overcome shortage of data; however, it cannot account for all the variations that can occur in real data.

Survey of Deep Learning Based Biomedical Image Segmentation Articles

In Table 18, a survey of recent articles involving deep learning approaches to biomedical image segmentation is provided. Only those articles which utilized deep learning models for biomedical image segmentation applications were selected. The table includes the article

reference, modality that defines the imaging techniques used for image formation or acquisition, methodology that defines the deep learning architecture used for segmentation, remarks section that briefly points out the proposed approach and finally the last column provides the performance metrics used for evaluating the proposed algorithm with brief results.

As seen from the Table 20, most of these approaches are based on either CNN or FCN based approaches. None of these articles utilized transfer learning approach while one article did use deep learning model as a function for feature extraction for subsequent classification with structured support vector machine. The modality for most of these applications were CT, MR and US which indicates the current trend of research as well. One reason for this is the ease of availability of image datasets through various competitions or from other public sources.

Table 18: Overview of Papers Using Deep Learning Techniques for Biomedical Image Segmentation				
Reference	Modality	Method	Remarks	Performance Metrics and Results
Badea et al. (2016) [87]	Clinical Images	CNN (LeNet and NiN)	LeNet and Network in Network (NiN) models were used for classification of burn images and for performance evaluation by comparing the classification accuracy for Skin vs Burn and Skin vs Light Burn vs Serious Burn.	Accuracy; LeNet was able to achieve an accuracy of 75.91% and 58.01% for classification of Skin vs Burn and Skin vs Light Burn vs Serious Burn, respectively. NiN achieved an accuracy of 55.7% for classification of Skin vs Light Burn vs Serious Burn.
Dhungel et al. (2015) [88]	X-Ray (Mammogram)	RBMs and CNN	Deep convolution and deep belief networks which are a type of RBM network are utilized as functions for conditional random field (CRF), and structured support vector machine (SSVM). The techniques are explored for segmenting breast masses from mammograms.	DICE; The Dice index of the proposed approach with all potential functions was 93% using CRF and 95% using SSVM.
Zhou et al. (2018) [69]	CT	CNN	Performance analysis of a proposed segmentation	Mean Accuracy, JSI;

			technique was performed for multiple organ detection.	The proposed approach achieved a mean JSI value of 79% and 67% for segmentation using 3D- and 2D deep CNN, respectively. Results are averaged for 17 types of organs.
Roth et al. (2018) [68]	CT	FCN (3D U-Net)	Deep 3D FCNs was used to automatically segment abdominal CT to delineate the arteries, portal vein, liver, spleen, stomach, gallbladder, and pancreas in each multi-organ image.	DICE; The proposed model achieved an average DICE score performance of (89.3 ± 6.5) % during testing.
Moeskops et al. (2016) [89]	MRI and CTA	CNN	Single CNN is used to perform multiple image segmentation tasks which includes six tissues in MR brain images, the pectoral muscle in MR breast images, and the coronary arteries in cardiac CTA. CNN identifies the imaging modality, anatomical structure and tissue class.	DICE; The proposed approach using a single CNN trained with three tasks had median percentage of voxels per scan below 0.0005% for all tasks when labelling a class which is alien to the target. This means that confusion between tasks was very low during labelling. An example of such a scenario is labelling of cortical grey matter in breast MR.
Dou et al. (2017) [90]	3D CT and 3D MRI	CNN	A 3D fully convolutional network combined with a 3D deep supervision mechanism is proposed for segmenting liver from 3D CT scans. Additionally, the approach is used for segmenting the whole heart and great vessels from 3D MR images.	JSI, DICE, Recall, Sensitivity, Specificity; The proposed approach for the evaluation of blood pool in heart segmentation task achieved JSI and DICE score of 86.5 and 92.8 respectively. This score was the best among participating teams. For myocardium segmentation, the scores were less with JSI and DICE values equal to 73.9% and 59.1%.
Wang et al. (2018) [91]	MRI	CNN	Proposed approach performs 2D segmentation of multiple organs (placenta, fetal brain, fetal lungs, maternal kidneys) from fetal MR slices, by using two organ annotations for training. Additionally, 3D segmentation of brain tumor core (without edema) and whole brain tumor (with edema) from MR sequences is performed using annotated	DICE; The proposed Bounding box and Image-specific Fine-tuning-based Segmentation (BIFSeg) achieved the best DICE score for placenta, fetal brain, fetal lungs, maternal kidneys with values of 86.41%, 90.39%, 85.35% and 86.33% respectively.

			tumor core in only one MR sequence for training.	
Havaei et al. (2017) [92]	MRI	CNN	An automatic brain tumor segmentation method based on multiple variations of CNN architecture was proposed. The real data was acquired from the 2013 brain tumor segmentation challenge (BRATS2013), which was part of the MICCAI conference. The dataset consisted of three subsets for training, testing and the leaderboard dataset for the competition containing 30, 10 and 25 patients respectively.	DICE, Sensitivity, Specificity; The TwoPathCNN achieved a DICE score, sensitivity and specificity of 85%, 93%, and 80% respectively. Out of the three cascade CNN architectures the InputCascadeCNN performed the best with a DICE score, sensitivity and specificity of 88%, 89%, and 87% respectively.
Ngo et al. (2017) [93]	MRI	RBMs	A segmentation approach is proposed for the endocardial and epicardial borders of the left ventricle (LV). The segmentation was done from all the slices of the end diastole (ED) and end systole (ES) cardiac phases of an MR cine study. The ED and ES volumes were manually selected by the user.	DICE; The proposed approach achieved an average DICE score of 90% compared with the boosted cascade detector's score of 86% for testing data.
Chen et al. (2018) [79]	3D MRI	3D ResNet	A voxel-wise residual network (VoxResNet) built with 25 layers is proposed for segmentation of key brain tissues into white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF).	DICE coefficient (DC), the 95th-percentile of the Hausdorff distance (HD) and absolute volume difference (AVD) The Dice coefficients results obtained for the proposed model for GM, WM, and CSF were 86.15%, 89.46%, and 84.25%, respectively when trained on relatively small training data.
Milletari et al. (2017) [94]	MRI and US	CNN and V-Net (FCN)	Six variants of CNN architectures were proposed and trained with patches derived from annotated medical volumes from MRI and transcranial US volumes illustrating respectively 26 areas of the basal ganglia and the midbrain. The proposed approach used parametric rectified linear units (PReLU) as activation functions where	DICE; The best DICE results for Hough-CNN was 85% compared with V-Net resulting in 71%.

			$PReLU\ (x) = \begin{cases} x\ if\ x \geq 0 \\ \alpha x\ if\ x < 0 \end{cases}$ The parameter α was learnt during training. Hough voting-based strategy was used to localize the anatomy of interest achieving high precision, despite the very high rate of mis-classified voxels.	
Xu et al. (2019) [78]	3D US	CNN	Proposed approach segments ultrasound images of breast into four major tissues: skin, fibro glandular tissue, mass, and fatty tissue using 3 orthogonal image planes; CNN indicates the tissue class of the centered pixel within the image blocks.	Accuracy, JSI, Precision, Recall, $F1_{measure}$; Above 80% performance results on these metrics.
Prince et al. (2019) [95]	Endoscopic OCT	CNN	An approach is proposed for automatically segmenting endoscopic OCT images using parallel architecture of trained deep neural network.	Accuracy; The proposed approach had an average relative error of prediction about 6.0% for five layers that included stratum corneum, epithelium, lamina propria, muscularis mucosae and submucosa.
Jia et al. (2017) [96]	Microscopic Images	FCN	An FCN based approach is proposed for image-to-image segmentation of histopathology images under deep weak supervision. Additionally, super-pixels instead of pixels were also used which is effective in maintaining intrinsic tissue boundaries.	$F1_{measure}$; The proposed algorithm was able to achieve the best $F1_{measure}$ of 83.6% which was significantly higher than existing weak supervised algorithms.
Zhao et al. (2018) [97]	Microscopic Images	Mask R-CNN	Model performs instance detection and instance segmentation on the nuclei of HL60 cells and microglia cells. For C.elegans embryo dataset because of scarcity of full voxel annotation only instance detection was performed.	Approximate Annotation Time (AT), $F1_{measure}$; Proposed approach was faster compared to VoxResNet with comparable performance. The average segmentation $F1_{measure}$ for the proposed approach and VoxResNet was 89.27% and 90.42% while AT was 5.5h and 22.5h respectively.

US – Ultrasound Images, OCT – Optical Coherence Tomography, MRI – Magnetic Resonance Imaging, CTA – Coronary Computed Tomography Angiogram

Discussion

In this section, a basic review of the deep learning approaches for some of the articles in Table 18 is provided.

Roth et al. proposed using multi-level deep convolutional networks for the segmentation of pancreas which is known to have very high anatomical variability. The segmentation of pancreas is important for quantifying organ volume for diabetic patients. Eighty-two contrast-enhanced abdominal CT volumes along with the ground truths were used in the study. Additionally, random non-rigid deformations were applied to obtain additional data instances. The degree of deformation was selected to ensure that the warped image had variations like the real data. In the proposed bottom-up approach, image patches were first labelled. For the image patch labelling, an axial, coronal, and sagittal view of the patches was used to obtain a per-location probability response map. This was followed by region labelling by generating super-pixel regions with high sensitivity but low precision at different spatial scales in a zoom out approach. CNN was used to assign probability to each super-pixel for being pancreatic tissue. Finally, the entire organ was detected from an abdominal CT scan using both the probability response map and CT intensity values obtained from the super-pixel regions.

Dou et al. proposed the use of a fully convolutional network in which all layers of the CNN were either convolutional or pooling for segmentation of liver from 3D CT scans, whole heart and great vessels from MR images. This approach was necessary to remove the size limits on the image input to the network which was there due to fully connected layers. The replacement of fully connected layers with convolutional layer implies that the input image can have any arbitrary size and classification output will be spatially arranged for the entire input image. This also eliminates the need of redundant computations resulting from overlapping regions in the traditional patch-

based approach. The 3D deep supervision mechanism simply extracts feature maps from the hidden layers, upscales them by connecting deconvolutional layers and using SoftMax function to obtain extra dense predictions. SoftMax function performs normalization on exponentials of input vector consisting of real numbers and converts them into probability distribution such that the output values are in the interval (0,1) with the sum of output values equals to 1. The classification error of these branch outputs in comparison with the ground truth was used to adjust network parameters of the mainstream network during training.

Milletari et al. proposed a patch-wise multi-atlas method by combining Hough based voting technique with CNN for segmenting regions of deep brain in MRI and ultrasound images. The approach can locally segment structures even if partially visible or corrupted by artifacts. The CNN is trained to classify foreground and background regions in the patches extracted from the voxels and links each input to its feature representation extracted from second last fully connected layer. The process is repeated for the entire training data and a dataset is generated that contains 2D, 2.5D or 3D patches of foreground regions and features extracted earlier. 2.5D contains higher spatial information of neighboring pixels and is typically obtained using three orthogonal patches (XY, YZ and XZ plane) with the resultant kernel still in 2D. Additionally, a vote is collected which is a displacement vector linking the voxel and position anatomy centroid. This is the same voxel from which the patch was collected. During testing, the CNN will generate the label for an unknown voxel and the corresponding feature map for all patches labeled as foreground will be extracted. Each feature map is compared with the feature maps in the database to extract k-closest neighbors based on Euclidean distance. Then the votes of these neighbors and corresponding segmentation patches are used to perform segmentation and the process is repeated for all foreground patches. This method was superior to voxel-wise semantic segmentation of CNNs in

111

all the parameter settings tested, required fewer training data, generated better segmentation contours, and eliminated the requirement for post-processing.

Chen et al. proposed voxel based deep learning model based on residual network (ResNet) and called as VoxResNet. Typically, deep learning models form feature representation in a successive manner with its level changes as low-middle-high. Greater the number of layers that are in the network means more information may be learned thus improving the discrimination capability of the network. But this does not always happen and at times performance starts to degrade as the network goes deeper. This is known as the degradation problem and ResNet have been shown to overcome this problem as the residual learning helps network optimization become easier. This is achieved through using skip connections. As a result, information can spread through the whole network both in forward and backward passes. In order to deal with volumetric data for brain segmentation from 3D magnetic resonance (MR) images, VoxResNet was proposed that extends 2D ResNet to 3D ResNet and its architecture is shown in Figure 27. As seen from the Figure 27, the proposed network consists of stacked residual modules. In VoxRes module, the input and transformed features are added together with skip connection. To deal with the size variations of 3D anatomical brain structures, 4 auxiliary classifiers C1-C4 are used for fusing multi-level contextual information. In the first layer of the network, information from multiple modalities that provided complimentary information regarding the same brain structure were combined with weights during training phase. This resulted in improved efficiency compared to results obtained through a single modality.
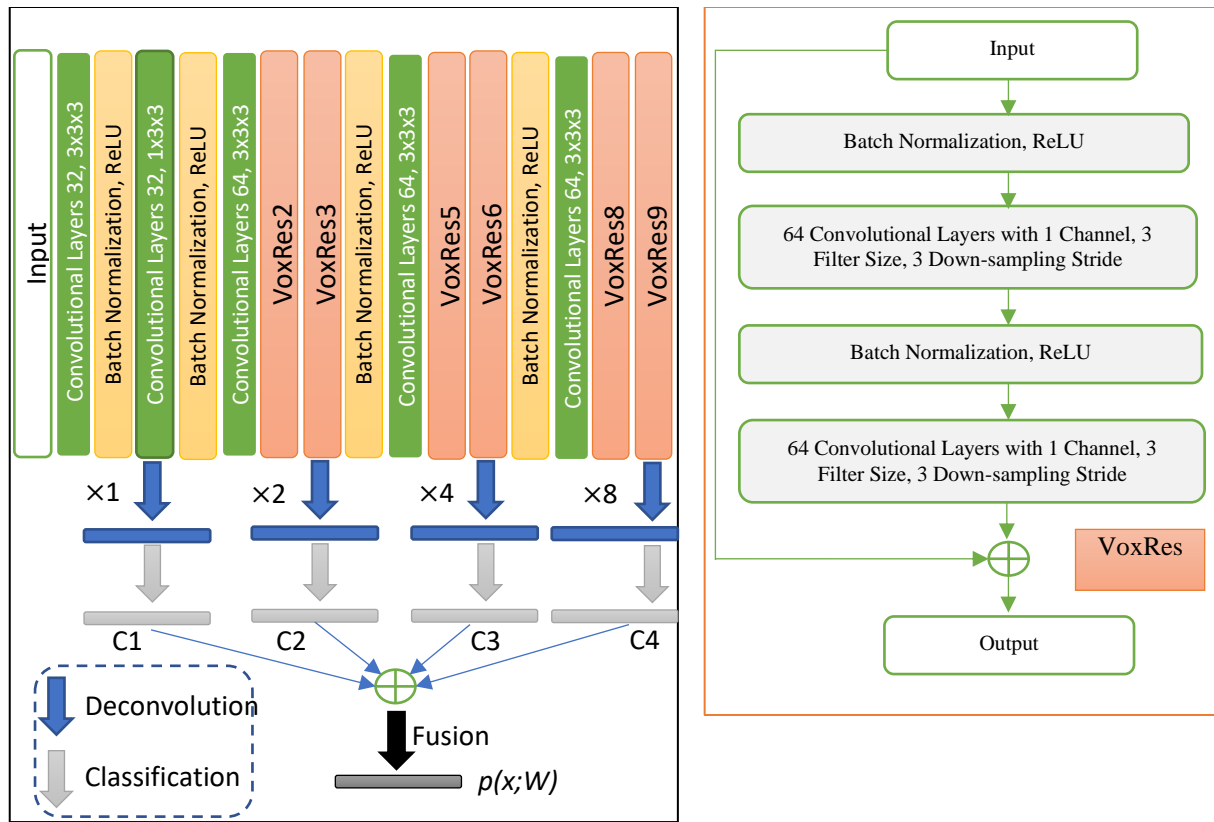
Figure 27. VoxResNet architecture for volumetric image segmentation. Figure shows batch normalization layers (BN), rectified linear units (ReLU), and convolutional layers N with number of channels, filter size and down-sampling stride. Figure adapted from Ref. [79].

Jia et al. proposed the use of deep weak supervision (DWS) under multiple instance learning (MIL) framework. In MIL approach during training of the classifier instances are grouped together and are called bags. Each bag is assigned a positive or negative label during the training, but the instances are not assigned any labels. The MIL approach requires that the classifier model should be able to predict not only instance level label but also has to perform bag level classification. From the perspective of histopathology images, each cancer or non-cancer image can be considered a bag while each pixel inside the image is the instance. Typically, a classifier predicts pixel level labels than image level prediction is evaluated using a softmax function or normalized exponential function. The DWS approach uses multiple side outputs taken from the CNN typically

after the convolutional layer similar to the way information was extracted by Chen et al. for fusion. The goal of DWS is to ensure minimal prediction error between ground truth and each of the side truths which results in improved performance. The two approaches were combined to form a new framework known as DWS-MIL. Additionally, another problem that occurs at times is the jagged tissue boundaries that results from treating pixels as instances. Use of super-pixels as instances was proposed to resolve this issue and decreases the number of instances thus reducing computational complexity.

Havaei et al. analyzed different custom architectures of CNN by concatenation of feature maps from different layers for the purpose of Brain tumor segmentation. In the first approach a local and global pathway was established to analyze changes in prediction accuracy due to visual details of the region surrounding the pixel and contextually in terms of the patch (containing the pixel) location in the brain. The architecture was called TwoPathCNN. The second architecture was based on cascading of two CNNs. The output layer of the first CNN was concatenated to one of the layers of the second CNN. Three different locations of second CNN were used for concatenation: 1) input layer (InputCascadeCNN), 2) first hidden layer in the second CNN (LocalCascadeCNN) and 3) right before the output layer of second CNN (MFCascadeCNN). These added inputs to the second CNN were made to analyze the impact of nearby labels to the final segmentation prediction. The two pathways model combined with collective training of the two convolutional pathways and with two training phases provided better segmentation compared to the single path architecture. The InputCascadeCNN as shown in Figure 28 outperformed the technique proposed by the winner of the BRATS 2013 challenge. Despite the use of two CNNs the segmentation time for the entire brain varied between 25 seconds to 3 minutes which was due to implementation of the architecture on graphical processing unit (GPU). The segmentation was

performed on the images in a slice-by-slice manner from the axial view with model sequentially processing each 2D slice.
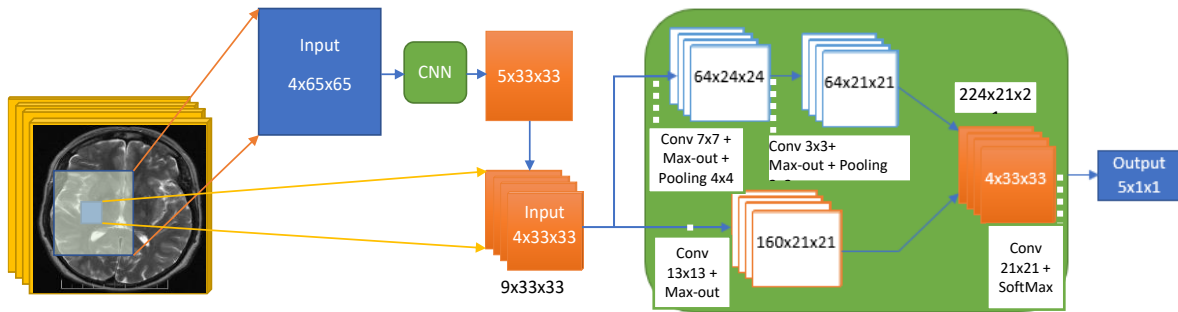

Figure 28. InputCascadeCNN, where 4x65x65 refers to number of input modalities x spatial width x spatial height. Adapted from Ref. [92].

Ngo et al. used Deep Belief Network (DBN) which is a type of RBMs and combined it with active contour model to utilize lower training data to achieve superior performance. The proposed approach utilized two separate DBNs for segmentation of endocardium and epicardium images despite similar appearance as it achieved high accuracy. Semi-automatic approach provided better results compared to fully automatic approach due to lack of enough training and testing data. Larger training data is required to design more complex DBN models. The existing image database which was part of MICCAI 2009 Left Ventricle (LV) segmentation challenge contained 15 sequences (from cardiac cine magnetic resonance) each for training, testing and online. Online dataset was originally made available on the challenge day for assessment of segmentation algorithms submitted by the participants. However, now the entire database is available along with the results achieved by the participants.

Badea et al. used LeNet CNN architecture, which was originally created in 1998 for handwritten digit recognition on bank notes, and Network in Network (NiN) architecture for classification of burn images. LeNet CNN comprises of 7 layers with three convolutional layers,

115

two pooling layers and one fully connected layer. The presence of a multilayer perceptron between the two main layers of a block differentiates NiN from LeNeT. Multilayer perceptron acts as a nonlinear function approximator to enhance the network's abstraction capability. The burn image database comprised of 611 images from 53 pediatric patients with image resolution of 1664 x 1248 pixels. The images were manually cropped to a final size of 230 x 240 pixels. Results indicated that the simple architecture performed well for binary classification problems, but performance significantly degraded with increasing complexity.

Zhao et al. demonstrated the use of deep learning model for the purpose of instance segmentation for 3D images to generate fully annotated data using a weakly supervised method [97]. Annotation of large number of biomedical images remains a challenge due to extensive time and labor requirements. The novelty of the proposed approach is that it requires a bounding box for all instances but only need a few full voxel annotations to achieve faster response and performance at par with the approach proposed by Chen et al. [79] that requires the full set of voxel annotation data. The two-step approach first detects all instances using 3D bounding box annotation and subsequently all detected instances are segmented. Mask R-CNN was used for 2D instance segmentation. Mask R-CNN can detect objects in an image and simultaneously generate segmentation mask for each instance. The proposed technique was tested on three biomedical datasets: nuclei of HL60 cells, microglia cells (in-house), and C. elegans developing embryos.

Chapter Conclusion

The review of deep learning approaches for biomedical image segmentation highlighted some key points. All these studies were based on empirical results that demonstrated the effectiveness of the proposed approach for the given application with limited datasets. The question that remains

is that why the deep learning approaches work for a given problem. The understanding of the answer to this question is an open area of research. Many researchers are working to develop novel visual approaches to help intuitive understanding of the feature maps obtained from the hidden layers [72], [98], [99]. Additionally, many researchers do not address the problem of generalizability of the network response in case the source of data changes. That is what will be the impact of change in data acquisition device as this may lead to changes in image characteristics such as illumination or color intensity levels. The lack of generalizability will have a negative impact on the network performance.

In the next chapter, the deep learning approaches for psoriasis image segmentation will be presented.

DEEP LEARNING BASED APPROACH FOR PSORIASIS IMAGE SEGMENTATION

In this chapter, a deep learning-based approach is proposed for the segmentation of psoriasis lesions using the digital images. As a first step, a digital image is acquired using a digital camera. Then, the background region is removed and only the portion of the image containing the skin regions are kept. To automatically identify the skin regions and the background region, an initial k-mean clustering is applied. The results are shown in Figure 29 below.

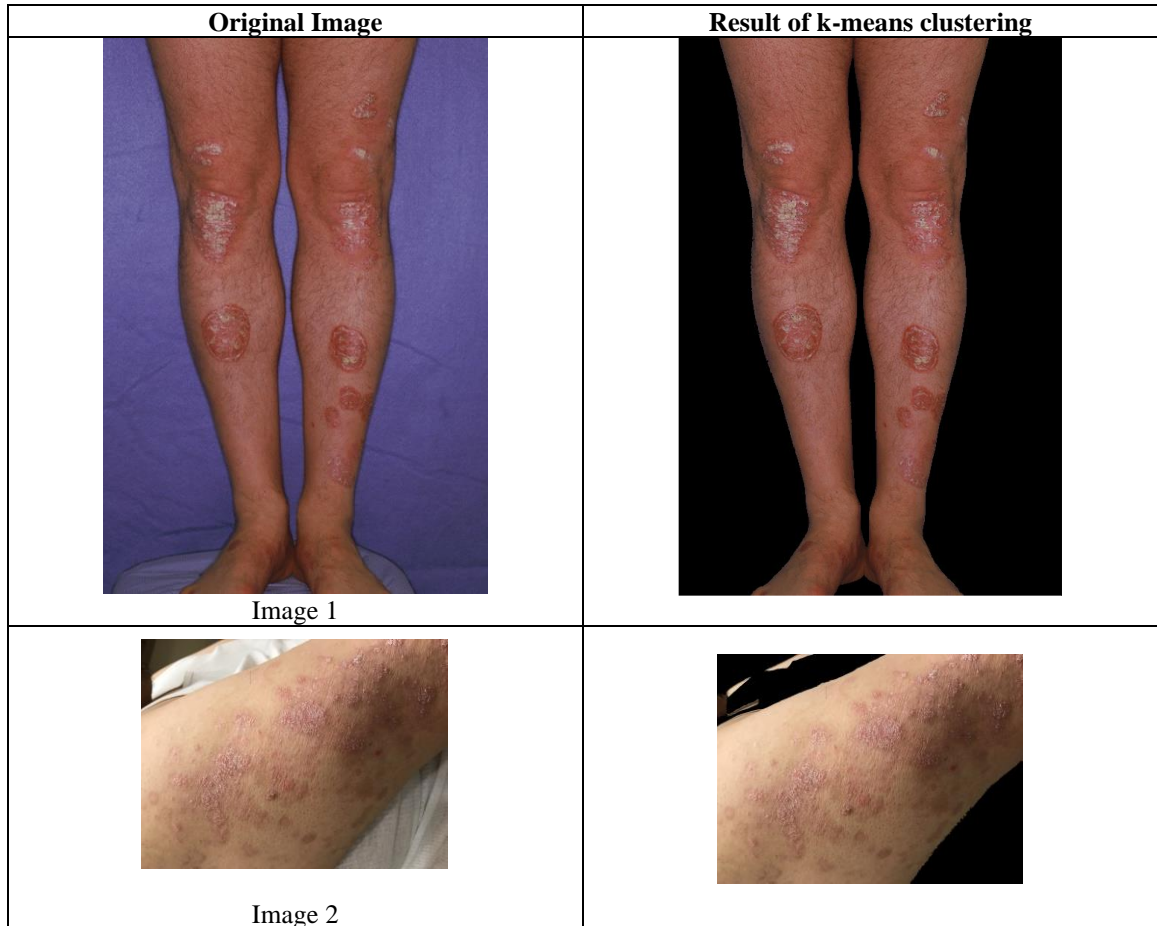| Original Image	Result of k-means clustering
Image 1	
Image 2	

Figure 29. Removal of Background Image

In the next step, images are further sub-divided into smaller patches containing the skin and lesion regions using superpixels based approach proposed in [59]. The superpixels approach applied on Image 1 is shown in Figure 30.
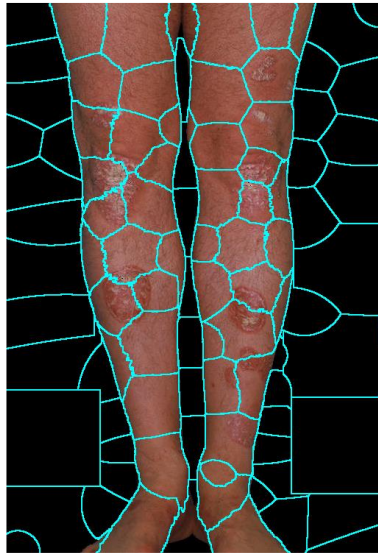

Figure 30. Division of Skin Regions into Small Images

Superpixels with mean value zero (black regions) can be discarded and remaining superpixels can be analyzed further for persence of lesions using semantic segmentation. Superpixels were used as an alternative to the fixed rectangular windows, as it has the added advantage of being dependent on pixel characteristics and can adapt its shape according to irregular lesion border which was not possible if a rectangular window was used. This can be seen from the Figure 31.

Image 1	Image 2

Figure 31. Superpixels Based Division of Skin Regions

Image 1 in Figure 31 shows how the superpixel boundary surrounds the lesions rather than cutting into the lesion. For very large lesions, the lesion will still be distributed into two regions as shown in Image 2 in Figure 31. For semantic segmentation of digital images containing normal skin and psoriasis lesions, a custom U-Net based deep neural network (DNN) is proposed. As mentioned in the last chapter, semantic segmentation involves classification of class labels for each pixel inside the image.

An image datastore was initially created which allows the file paths to be specified for the images without having to load all the images in the dataset in the memory. This allows for use of large number of images without running into out of memory issues during execution. The image dataset was the same as used earlier for the evaluation of proposed clustering algorithms. As mentioned earlier, the resolution of the images in the dataset varies from 154 by 151 to 1800 by 1700 pixels. To use DNN, the images must be resized to a uniform size. For the custom U-Net network an image size of [224 330 3] was selected. This size was selected to balance the issues arising from information loss due to downscaling of high-resolution images and addition of noisy pixels due to up-scaling of low-resolution images. Bilinear interpolation was used for image resizing in which the output pixel value is determined using the weighted average of pixels in the nearest 2-by-2 neighborhood. Bicubic interpolation uses a larger neighborhood of 4-by-4 pixels but was not used as the resulting output pixel values may lie outside of the original range of pixel values.

The resized images were saved in bitmap (BMP) format instead of Joint Photographic Experts Group (JPEG) format. BMP was preferred as it stores color data with each pixel of the image without any compression providing highest quality of the images but resulting in a larger storage size for each image. JPEG on the other hand applies compression on the color data associated with

the image pixels resulting in loss of information that maybe required for subsequent DNN Model training and testing. Apart from image dimension resizing, contrast enhancement was done using histogram equalization. The Figure 32 provides resized images before and after contrast enhancement.

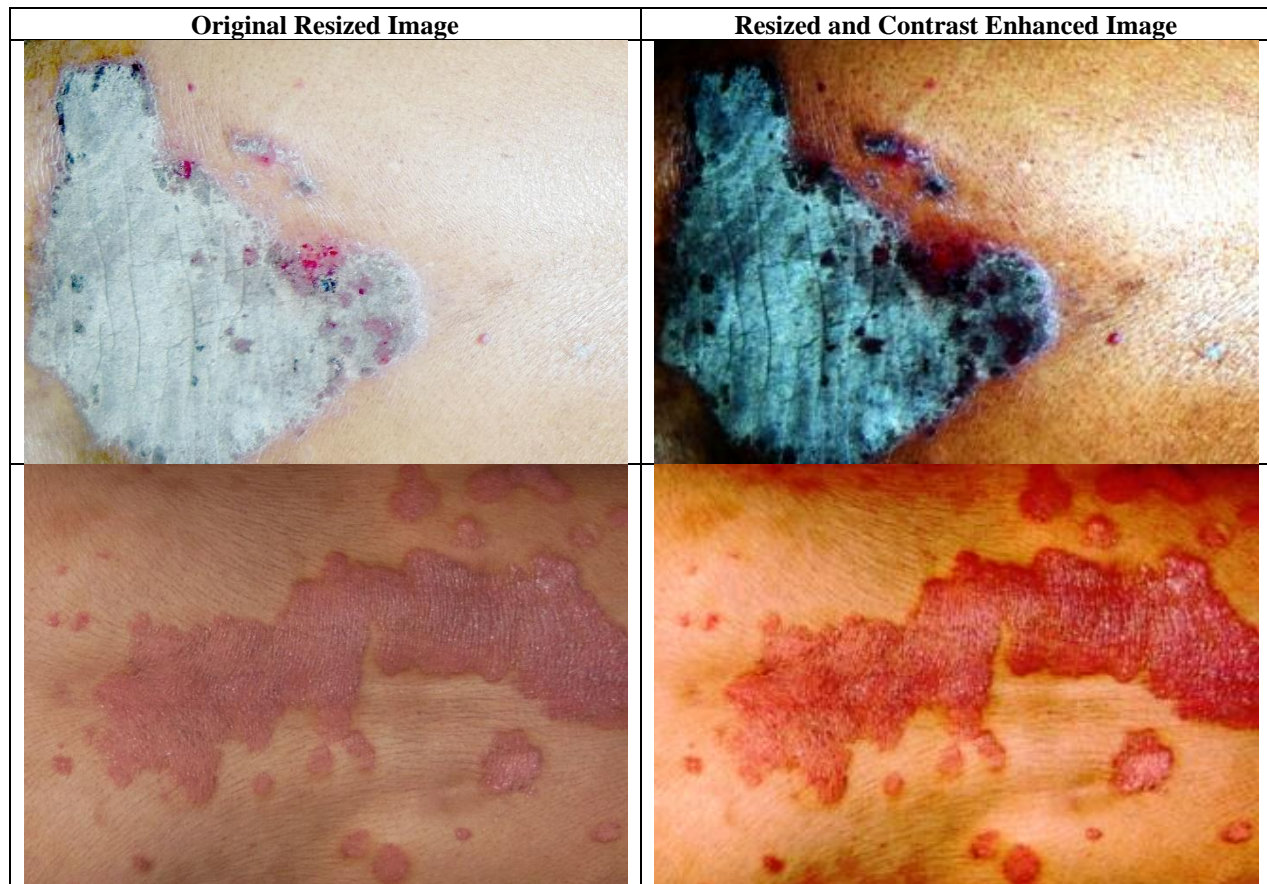| Original Resized Image	Resized and Contrast Enhanced Image

Figure 32. Effect of Contrast Enhancement

After the basic pre-processing of the images, dataset statistics were analyzed. Specifically, the distribution of class labels was determined by first counting the number of pixels associated with each class label and then taking its ratio with the total number of pixels in the dataset. Mathematically, $Frequency = \frac{count\ of\ pixels\ in\ each\ class}{Total\ number\ of\ pixels}$. The results are shown in Figure 33. As seen from the distribution, the normal skin pixels are higher than the psoriasis lesion and this is typically true for most of the digital images in the dataset. The class imbalance must be adjusted

to ensure the DNN does not produce bias results as the learning process may produce more positive results for the dominant class.
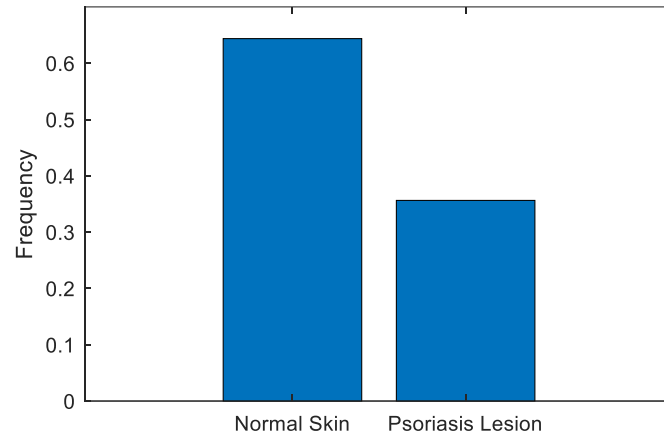

Figure 33: Pixel distribution of normal skin and lesion region

This class imbalance is addressed by determining the median frequency class weights using the equation: $classWeights = \frac{median(Frequency)}{Frequency}$. For the current dataset, the class weights were determined to be [0.7768, 1.4032]. As seen, the class weights associated with the dominant class was adjusted to be a lower value compared to the minor class.

In the next step, the 80% of the dataset was randomly distributed for training the DNN while the remaining 20% was used for testing and validating the results of the training. That is 36 images were used for training and 9 images were used for testing.

After data distribution, U-Net DNN was created. U-Net was determined to be the best network for this application after experimenting with different base networks such as ResNet-50 and MobileNet v2 with both networks' segmentation accuracy was below 70%. Other semantic segmentation network architectures such as SegNet and fully convolutional networks (FCN) were also tested, and their accuracy remained below 80%. U-Net is based on encoder-decoder based

architecture and since it does not have any dense layer it can accept images of any size. For the current design an encoder depth of 4, was chosen. Encoder Depth (D) specifies the number of times the input image is down sampled or up sampled during processing. The encoder-decoder network samples the image by a factor of 2^D. where D is the value of Encoder Depth. The created network is shown in Figure 34.
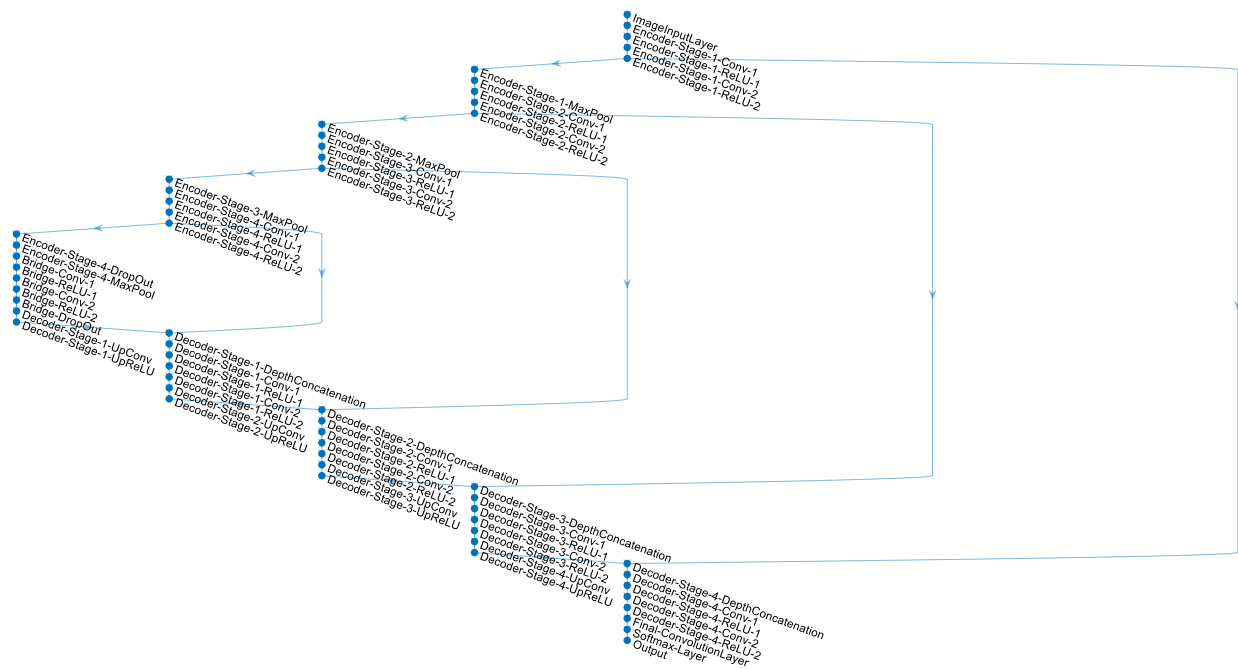

Figure 34. U-Net architecture

As seen from the Figure 34, the input layer is at the top and the final output layer is at the bottom. The upper half of the figure is the encoder, and the lower half of the figure is the decoder. The closer inspection of the network shows that each layer contains traditional convolutional and max pooling layers. As mentioned earlier, down sampling is performed in the encoder part of the network design and is achieved through convolutional layer that extracts the context of the image by iteratively moving across the image over a small region while max pooling reduces the size of feature map for achieving lower network complexity by retaining fewer parameters. To elaborate

this further, as an example we consider that the convolutional layer forms small regions of size 4x4 while max pooling layer further reduces this size to 2x2 while retaining all the essential information.

After the network is created, data augmentation was used for increasing the number of training examples as the training set is small. For data augmentation following techniques were used: random reflection in the left-right direction, random reflection in the top-bottom direction, image rotation using the rotation angle picked randomly from a continuous uniform distribution within the specified interval, horizontal scaling using the horizontal scale factor picked randomly from a continuous uniform distribution within the specified interval, vertical scaling, horizontal shear, vertical shear, horizontal translation and vertical translation.

The next step was to train the network. The network training results are shown in Figure 35.
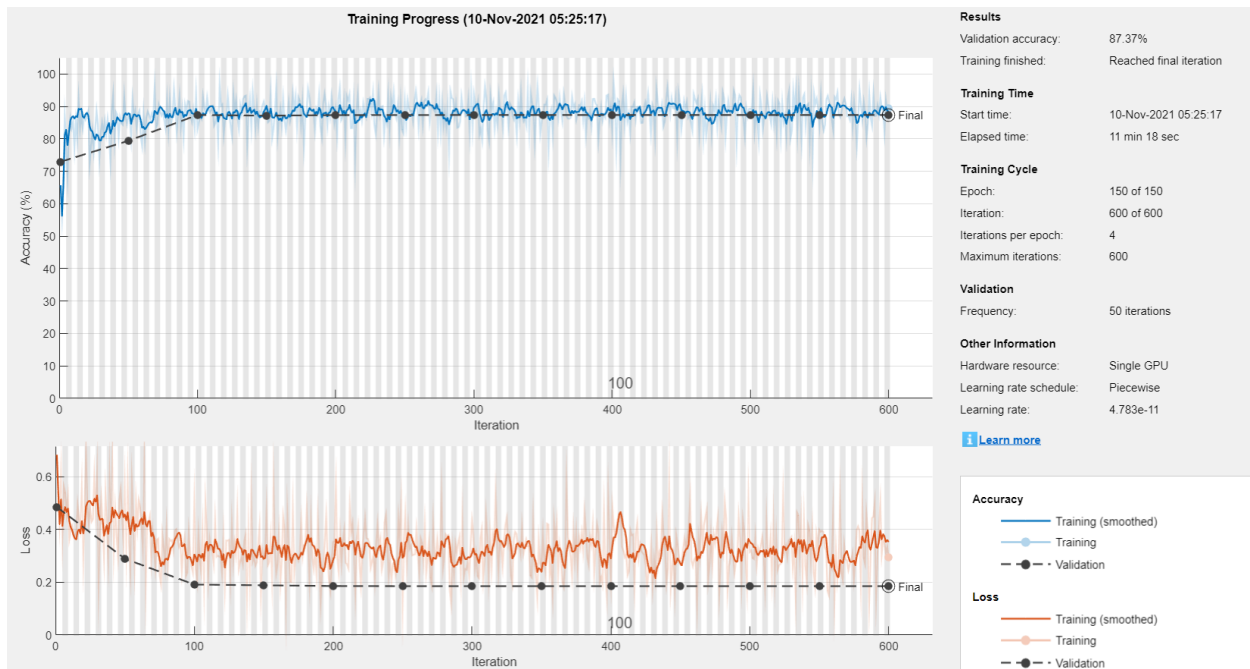
Figure 35: Training and Validation Results

As seen from the Figure 35, the network was able to achieve an overall validation accuracy of

87.37%. The Intersection over Union results of the network is provided in Table 19.

Table 19: Intersection over Union (IoU) Results	
Class Label	IoU
Normal Skin	0.8979
Psoriasis Lesion	0.8796

Results for some of the sample images are provided below.

Original Image	Segmented Image
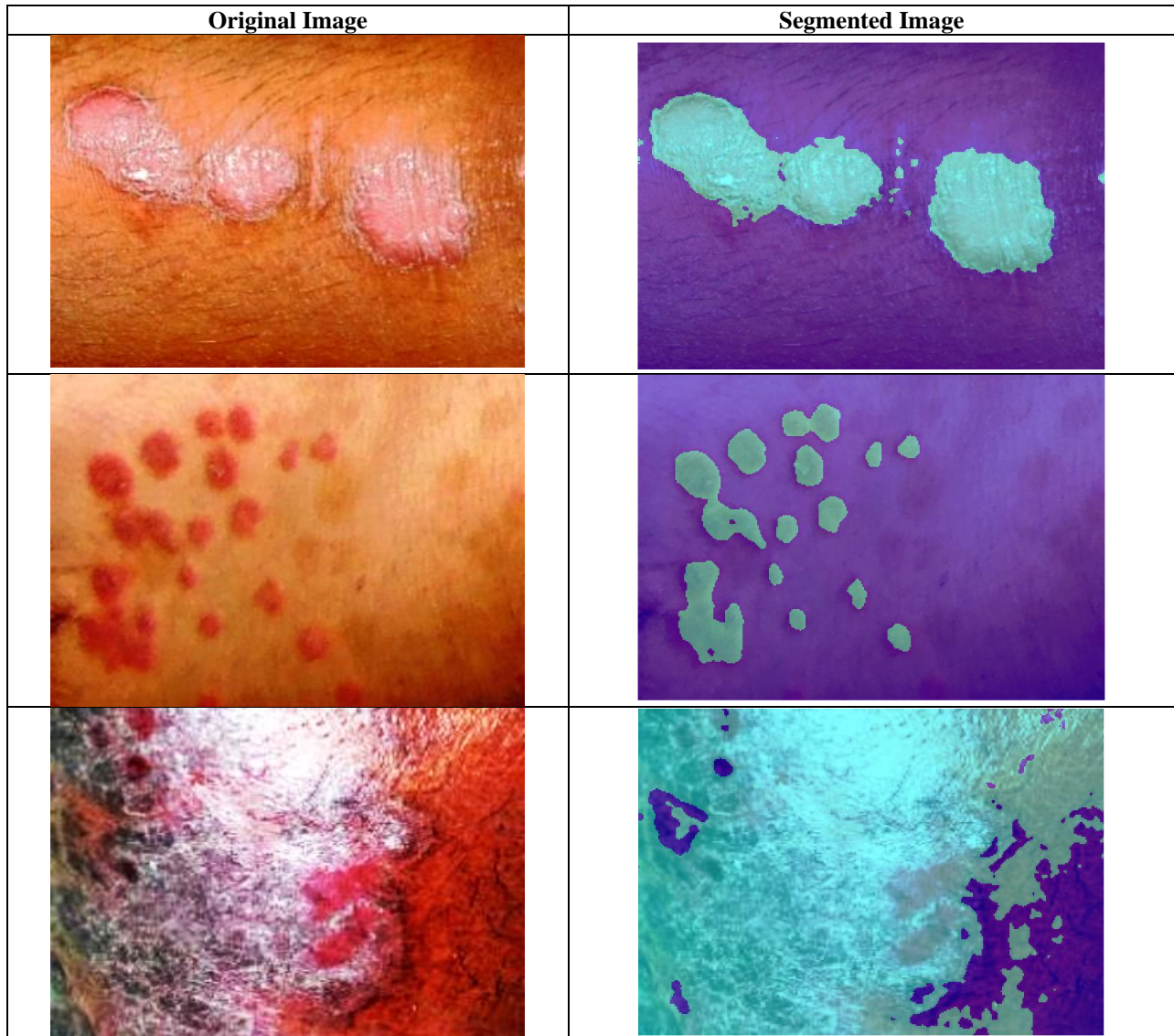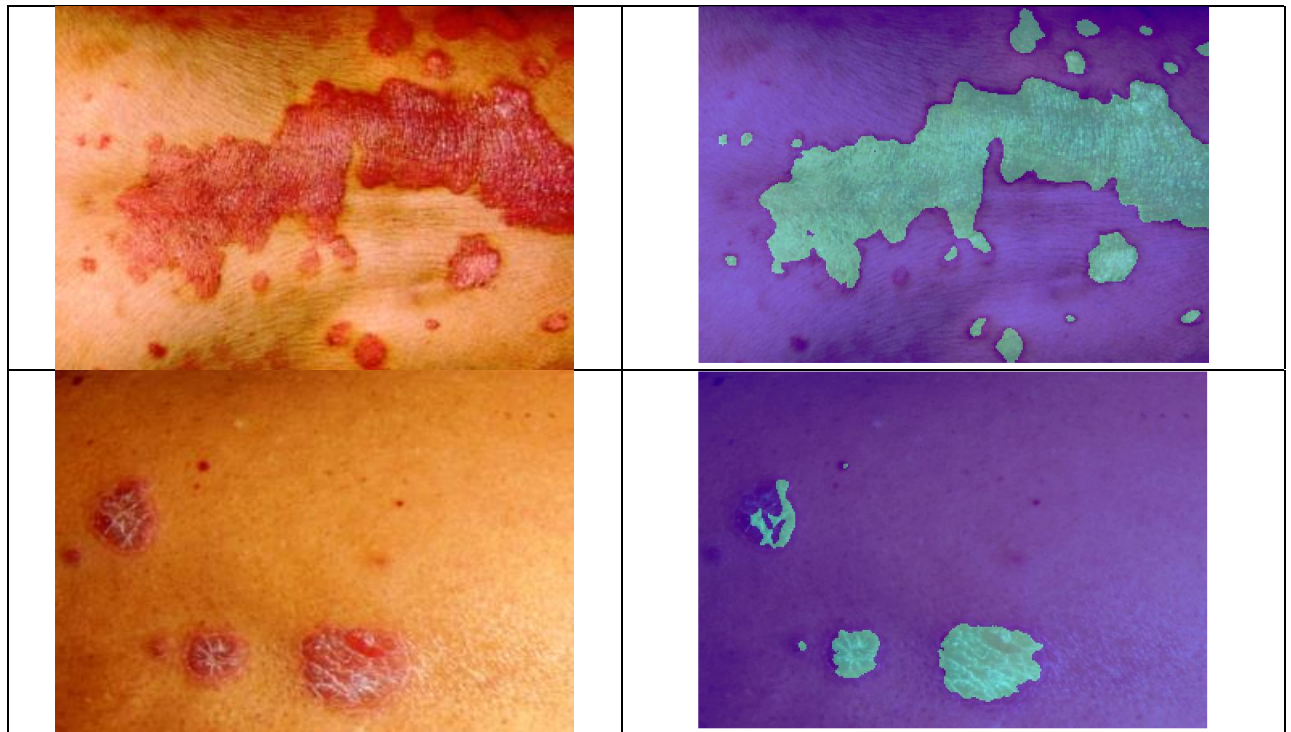
Figure 36. Visual Examination of Segmentation Results

As seen from the above images, most lesions were accurately identified by the network. These results were obtained with minimal pre and post processing of the images. To validate the trained DNN further, it was tested using another set of digital images collected from the internet. Results on some of the images are provided in the Figure 37 below.

Original Image	Segmented Image
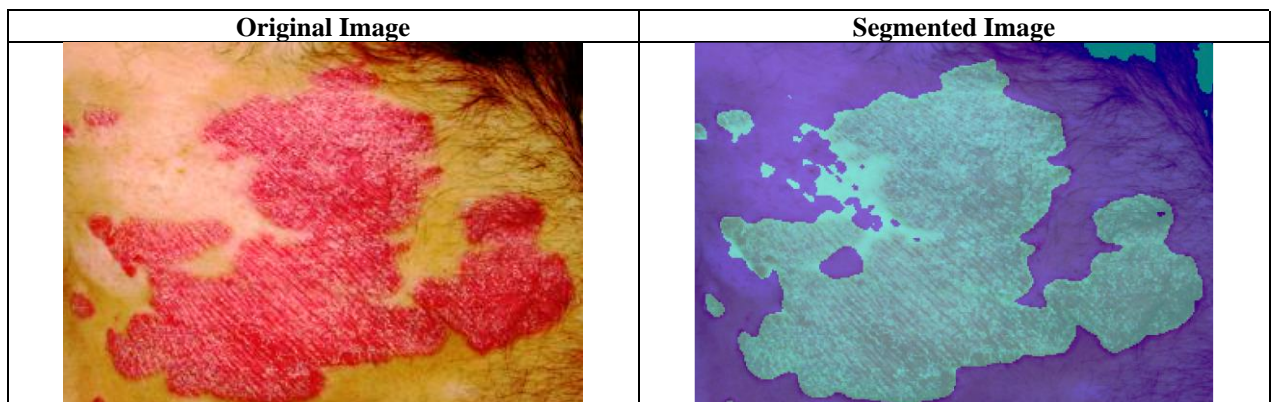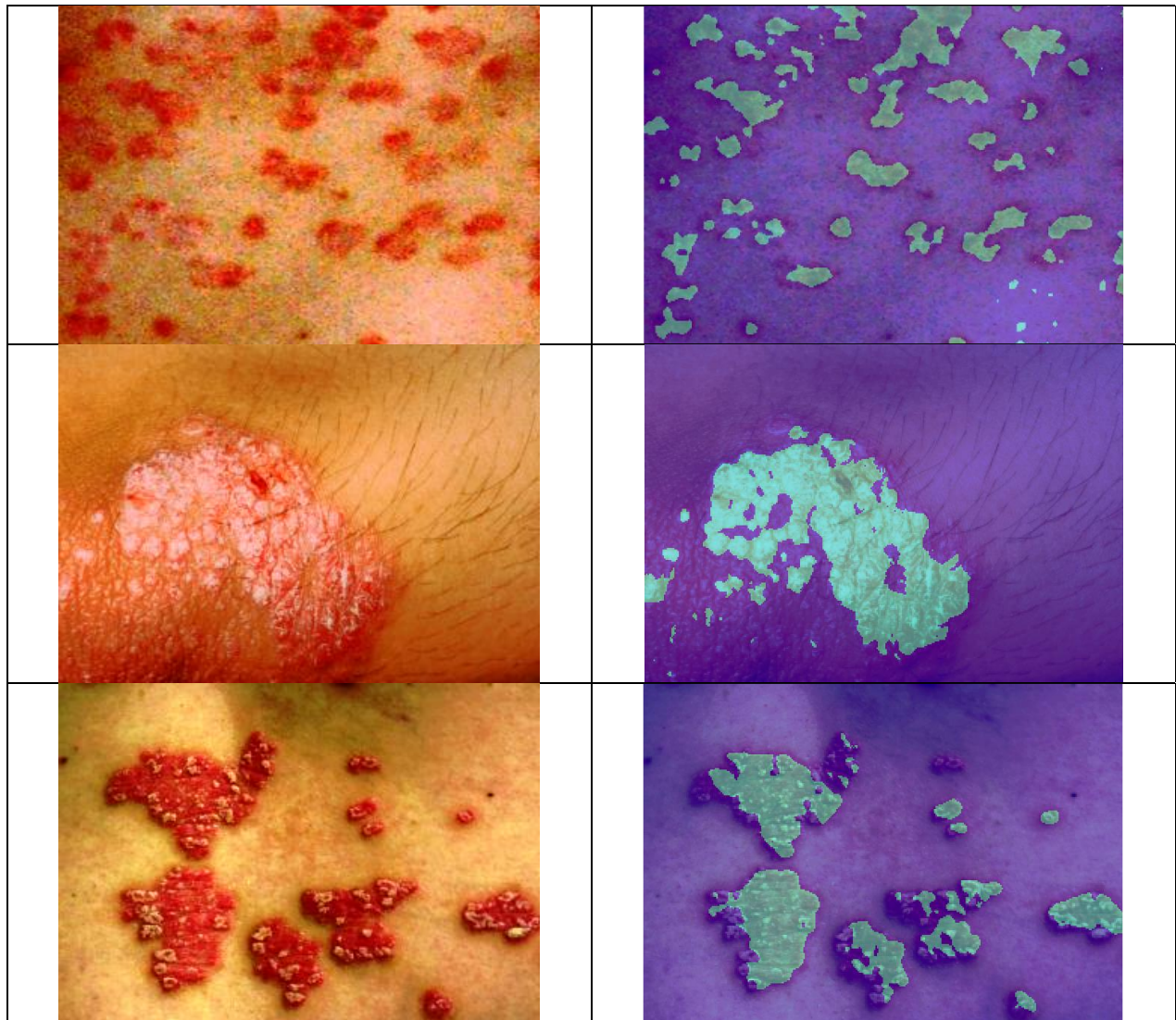
Figure 37. Segmentation Results on New Image Dataset

The images shown in Figure 37, were collected from the internet to assess the performance of the trained network on a new set of images which the network has not previously seen. This set can be referred to as a blind set. This allows us to test the trained network performance when the image acquisition device, light conditions, distance of the camera from the lesion and resolution of the images are not under our control. The selection criteria for images in the blind set was the visual similarity of the selected image with the images already present in the existing dataset. The

network cannot be evaluated on a set of images which have very little similarity with the existing images. Because that will require, retraining of the network with additional images.

Chapter Conclusion

In this chapter, the various stages of the segmentation of the psoriasis lesions from image acquisition to lesion identification was presented based on DNN based approach using custom trained U-Net. The network was able to achieve an accuracy of ~87% with minimal pre and post processing. The approach outlined shows that even in the presence of small image dataset DNNs can be utilized to achieve reasonably high accuracies.

CONCLUSION

Based on the research presented in this dissertation, following major conclusions can be found:

- The analysis of the existing mobile apps, interviews with the psoriasis patients and dermatologists showed that there exists a clear need for an objective and an automatic evaluation of psoriasis disease severity. One of the major limitations of the study was the absence of a large image database which can be universally used to evaluate various machine learning and deep learning based approaches. The work presented in this dissertation, can be used to set up such a database by using these approaches in combination with visual examination to perform coarse segmentation on a large set of images. While human supervision can be used to fine tune some of the lesions that may have been missed or wrongly segmented by the algorithms.

- The work showed that traditional machine learning based approaches to algorithm design achieve superior performance but remain time consuming to build and adapt under changing scenarios. Multiple different algorithms were proposed using traditional clustering approaches which have the benefit of being unsupervised techniques and achieved an average maximum accuracy of ~94% across the 45 images.

- The DNN based approach was able to achieve an accuracy of ~87% which was higher than the accuracy achieved through SVM based approach. The SVM based approach required intensive human feature engineering while DNN based approach required selection of a correct base network, hyperparameter optimizations of the network and selection of appropriate training parameters. The results did not require any post processing.

- Lastly, the classification of images into red, red + scaling and scaling class was proposed as an additional parameter to assess the psoriasis severity. The proposed equations allow

for the quantitative determination of redness level and presence of scales inside the lesion based on the color profile of the lesions. This provides an objective mechanism to assess the redness level of the lesion and severity of the scales as required by the PASI scale.

As part of the future work, it is imperative that dermatologists keep visual records of lesions as they change over time to keep track of disease severity. And the results obtained from the image analysis needs to be integrated with the Electronic Health Record (EHR) as proposed in the business canvas model. This will help in correlating this information with medication intake to determine treatment effectiveness and also enable tracking of flare-ups which in turn can be used to determine patient specific disease triggers.

REFERENCES

[1]   A. D. Parisi R, Symmons DPM, Griffiths CEM, "The identification and management of psoriasis and associated comorbidity project team. Global epidemiology of psoriasis: a systematic review of incidence and prevalence," *J. Invest. Dermatol.*, vol. 133, no. 2, pp. 377–385, 2013.

[2]   V. Chandran and S. P. Raychaudhuri, "Geoepidemiology and environmental factors of psoriasis and psoriatic arthritis," *J. Autoimmun.*, vol. 34, no. 3, pp. J314–J321, 2010, doi: 10.1016/j.jaut.2009.12.001.

[3]   "Psoriasis statistics," *MG217*, 2016. http://www.mg217.com/your-psoriasis/statistics-about-psoriasis/ (accessed May 06, 2016).

[4]   "Global reports on psoriasis," 2016. [Online]. Available: http://apps.who.int/iris/bitstream/10665/204417/1/9789241565189_eng.pdf.

[5]   A. Y. Marks, R., Barton, S. P., Shuttleworth, D., & Finlay, "Assessment of disease progress in psoriasis," *Arch. Dermatol.*, vol. 125, no. 2, pp. 235–240, 1989.

[6]   N. C. for C. D. P. and Centers for Disease Control and Prevention, H. Promotion, D. Control, C. D. Prevention, H. P. This, and T. Order, "Developing and Addressing the Public Health Agenda for Psoriasis and Psoriatic Arthritis," *Cdc*, 2010. https://www.cdc.gov/psoriasis/pdf/public-health-agenda-for-psoriasis.pdf (accessed Jun. 20, 2017).

[7]   B. Strober, C. Karki, M. Mason, J. D. Greenberg, and M. Lebwohl, "Impact of Psoriasis Area and Severity Index (PASI) on patient reported outcomes in patients with psoriasis : Results from the Corrona Psoriasis Registry," pp. 1–14, 2017.

[8]   A. Menter and E. Lilly, "Focus on Managing Patients with Moderate-to-Severe Plaque Psoriasis."

[9]   L. Mertz, "mHealth to the Recue: Growing use of wireless and mobile technologies improves community health, even in rural areas.," *IEEE Pulse*, vol. 7, no. 6, pp. 16–24, 2016, doi: 10.1109/MPUL.2016.2606468.

[10]   World Health Organization, "mHealth: New horizons for health through mobile technologies," *Observatory*, vol. 3, no. June, pp. 66–71, 2011, doi: 10.4258/hir.2012.18.3.231.

[11]   R.-G. Jahns and P. Houck, "Mobile Health Market Report 2013-2017," *Research2Guidance*, p. 116, 2013, [Online]. Available: http://www.research2guidance.com/shop/index.php/mhealth-app-performance-benchmarking.

[12]   Research2Guidance, "mHealth App Developer Economics Report 2014," *Research2Guidance*, pp. 1–43, 2014, doi: 10.15971/j.cnki.cmdi.2016.06.002.

[13]   P. Olla and C. Shimskey, "mHealth taxonomy: a literature survey of mobile health applications," *Health Technol. (Berl).*, vol. 4, no. 4, pp. 299–308, 2015, doi: 10.1007/s12553-014-0093-8.

[14]   Z. M. Mao, "Diagnosing mobile apps' quality of experience: Challenges and promising directions," *IEEE Internet Comput.*, vol. 20, no. 1, pp. 66–69, 2016, doi: 10.1109/MIC.2016.1.

[15]   M. J. Handel, "MHealth (mobile health)-Using Apps for health and wellness," *Explor. J. Sci. Heal.*, vol. 7, no. 4, pp. 256–261, 2011, doi: 10.1016/j.explore.2011.04.011.

[16]   S. Khoja, H. Durrani, R. E. Scott, A. Sajwani, and U. Piryani, "Conceptual Framework for Development of Comprehensive e-Health Evaluation Tool," *Telemed. e-Health*, vol. 19, no. 1, pp. 48–53, 2013, doi: 10.1089/tmj.2012.0073.

[17]   L. Shamseer *et al.*, "Preferred reporting items for systematic review and meta-analysis protocols (prisma-p) 2015: Elaboration and explanation," *BMJ*, vol. 349, no. January, pp. 1–25, 2015, doi: 10.1136/bmj.g7647.

[18]   S. S. Zaghloul, M. Jeremy, and D. Goodfield, "Objective Assessment of Compliance With Psoriasis Treatment," vol. 140, 2017.

[19]   C. L. Carroll, S. R. Feldman, F. T. Camacho, and R. Balkrishnan, "Better medication adherence results in greater improvement in severity of psoriasis," *Br. J. Dermatol.*, vol. 151, no. 4, pp. 895–897, 2004, doi: 10.1111/j.1365-2133.2004.06174.x.

[20]   U. Kingdom, "1 Patients with psoriasis and their compliance with medication," pp. 581–583.

[21]   N. Dey, A. S. Ashour, F. Shi, and R. S. Sherratt, "Wireless Capsule Gastrointestinal Endoscopy: Direction-of-Arrival Estimation Based Localization Survey," *IEEE Rev. Biomed. Eng.*, vol. 10, pp. 2–11, 2017, doi: 10.1109/RBME.2017.2697950.

[22]   W. J. van der Meer, F. S. Andriessen, D. Wismeijer, and Y. Ren, "Application of intra-oral dental scanners in the digital workflow of implantology," *PLoS One*, vol. 7, no. 8, pp. 1–8, 2012, doi: 10.1371/journal.pone.0043312.

[23]   H. D. S. Ganzeli, J. G. Bottesini, L. D. O. Paz, and M. F. S. Ribeiro, "SKAN: Skin Scanner - System for Skin Cancer Detection Using Adaptive Techniques," *IEEE Lat. Am. Trans.*, vol. 9, no. 2, pp. 206–212, 2011, doi: 10.1109/TLA.2011.5765575.

[24]   Shankaracharya, D. Odedra, S. Samanta, and A. S. Vidyarthi, "Computational intelligence in early diabetes diagnosis: A review," *Rev. Diabet. Stud.*, vol. 7, no. 4, pp. 252–261, 2010, doi: 10.1900/RDS.2010.7.252.

[25]   A. Loddo, C. Di Ruberto, and M. Kocher, "Recent advances of malaria parasites detection systems based on mathematical morphology," *Sensors (Switzerland)*, vol. 18, no. 2, pp. 1–21, 2018, doi: 10.3390/s18020513.

[26]   V. K. Shrivastava, N. D. Londhe, R. S. Sonawane, and J. S. Suri, "First review on psoriasis severity risk stratification: An engineering perspective," *Comput. Biol. Med.*, vol. 63, pp. 52–63, Aug. 2015, doi: 10.1016/j.compbiomed.2015.05.005.

[27]   P. I. Spuls, L. L. A. Lecluse, M.-L. N. F. Poulsen, J. D. Bos, R. S. Stern, and T. Nijsten, "How Good Are Clinical Severity and Outcome Measures for Psoriasis?: Quantitative Evaluation in a Systematic Review," *J. Invest. Dermatol.*, vol. 130, no. 4, pp. 933–943, 2010, doi: 10.1038/jid.2009.391.

[28]   A. Bożek and A. Reich, "The reliability of three psoriasis assessment tools: Psoriasis area and severity index, body surface area and physician global assessment," *Adv. Clin. Exp. Med.*, vol. 26, no. 5, pp. 851–856, 2017, doi: 10.17219/acem/69804.

[29]   S. R. Feldman and G. G. Krueger, "Psoriasis assessment tools in clinical trials," *Ann. Rheum. Dis.*, vol. 64, no. SUPPL. 2, pp. 65–69, 2005, doi: 10.1136/ard.2004.031237.

[30]   S. Lane *et al.*, "Assessing Severity In Psoriasis: Correlation Of Different Measures (Pasi, Bsa, And Iga) In A Canadian Real-World Setting," *Value Heal.*, vol. 19, no. 3, p. A122, 2016, doi: 10.1016/j.jval.2016.03.490.

[31]   J. A. Walsh, H. Tan, D. Ph, H. Valdez, and K. C. Duffin, "Comparative Assessment of PASI and Variations of PGA × BSA as Measures of Psoriasis Severity," no. 4, pp. 113–118, 2017, doi: 10.1177/247553031700200409.

[32]   V. Oji and T. A. Luger, "The skin in psoriasis: Assessment and challenges," *Clin. Exp. Rheumatol.*, vol. 33, pp. 14–19, 2015.

[33]   M. Boston, "Why Assess and Quantitate the Extent of both Psoriasis and Psoriatic Arthritis in Clinic ? DISCLOSURE OF RELEVANT."

[34]   M. Augustin *et al.*, "A framework for improving the quality of care for people with psoriasis," *J. Eur. Acad. Dermatology Venereol.*, vol. 26, pp. 1–16, 2012, doi: 10.1111/j.1468-3083.2012.04576.x.

[35]   T. T. K. Munia *et al.*, "Automatic clustering-based segmentation and plaque localization in psoriasis digital images," in *2017 IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT)*, Nov. 2017, pp. 113–116, doi: 10.1109/HIC.2017.8227597.

[36]   E. Tellier, "Medical-Image-Processing." https://github.com/EliseTellier/Medical-Image-Processing/commit/70449f6212cc9f481f554d02db5fc610324b0eee (accessed Aug. 30, 2017).

[37]   A. Pal, A. Roy, K. Sen, R. Chatterjee, U. Garain, and S. Senapati, "Mixture model based color clustering for psoriatic plaque segmentation," *Proc. - 3rd IAPR Asian Conf. Pattern Recognition, ACPR 2015*, pp. 376–380, 2015, doi: 10.1109/ACPR.2015.7486529.

[38]   F. Bogo, M. Samory, A. B. Fortina, S. Piaserico, and E. Peserico, "Psoriasis segmentation through chromatic regions and Geometric Active Contours," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2012, pp. 5388–5391, doi: 10.1109/EMBC.2012.6347212.

[39]   A. Pal, U. Garain, R. Chatterjee, and S. Senapati, "Psoriatic plaque segmentation in skin images," in *2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, Dec. 2015, pp. 1–4, doi: 10.1109/NCVPRIPG.2015.7489994.

[40]   J. S. Taur, G. H. Lee, C. W. Tao, C. C. Chen, and C. W. Yang, "Segmentation of psoriasis vulgaris images using multiresolution-based orthogonal subspace techniques," *IEEE Trans. Syst. Man, Cybern. Part B Cybern.*, vol. 36, no. 2, pp. 390–402, 2006, doi: 10.1109/TSMCB.2005.859935.

[41]   J. Lu, E. Kazmierczak, J. H. Manton, and R. Sinclair, "Automatic segmentation of scaling in 2-D psoriasis skin images," *IEEE Trans. Med. Imaging*, vol. 32, no. 4, pp. 719–730, 2013, doi: 10.1109/TMI.2012.2236349.

[42]   N. K. Al Abbadi, N. S. Dahir, and H. Restom, "Psoriasis Detection Using Skin Color and Texture Features," *J. Comput. Sci.*, vol. 6, no. 6, pp. 648–652, 2010.

[43]   M. N. Alam, T. T. K. Munia, K. Tavakolian, F. Vasefi, N. Mackinnon, and R. Fazel-Rezai, "Automatic detection and severity measurement of eczema using image processing," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2016-Octob, pp. 1365–1368, 2016, doi: 10.1109/EMBC.2016.7590961.

[44]   B. D. Baswaraj, A. Govardhan, and P. Premchand, "Active Contours and Image Segmentation: The Current State of the Art," vol. 12, no. 11, 2012.

[45]   V. K. Shrivastava, N. D. Londhe, R. S. Sonawane, and J. S. Suri, "Exploring the color feature power for psoriasis risk stratification and classification: A data mining paradigm," *Comput. Biol. Med.*, vol. 65, no. 2015, pp. 54–68, 2015, doi: 10.1016/j.compbiomed.2015.07.021.

[46]   Y. George, M. Aldeen, and R. Garnavi, "Pixel-based skin segmentation in psoriasis images," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug. 2016, vol. 2016-Octob, pp. 1352–1356, doi: 10.1109/EMBC.2016.7590958.

[47]   L. H. Juang and M. N. Wu, "Psoriasis image identification using k-means clustering with morphological processing," *Meas. J. Int. Meas. Confed.*, vol. 44, no. 5, pp. 895–905, 2011, doi: 10.1016/j.measurement.2011.02.006.

[48]   J. Glaister, R. Amelard, A. Wong, and D. A. Clausi, "MSIM: Multistage illumination modeling of dermatological photographs for illumination-corrected skin lesion analysis," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 7, pp. 1873–1883, 2013, doi: 10.1109/TBME.2013.2244596.

[49]   C. Barata, J. S. Marques, and M. E. Celebi, "Improving dermoscopy image analysis using color constancy," *2014 IEEE Int. Conf. Image Process.*, vol. 19, no. 3, pp. 3527–3531, 2014, doi: 10.1109/ICIP.2014.7025716.

[50]   D. Arthur and S. Vassilvitskii, "k-means ++ : The Advantages of Careful Seeding," vol. 8, pp. 1–11.

[51]   P. Soille, *Morphological Image Analysis: Principles and Applications*. Springer-Verlag, 1999.

[52]   T. T. K. Munia, M. N. Alam, J. Neubert, and R. Fazel-rezai, "Automatic Diagnosis of Melanoma Using Linear and Nonlinear Features from Digital Image," 2017.

[53]   K. H. Zou *et al.*, "Statistical Validation of Image Segmentation Quality Based on a Spatial Overlap Index," *Acad. Radiol.*, vol. 11, no. 2, pp. 178–189, 2004, doi: 10.1016/j.molcel.2007.05.041.A.

[54]   R. Achanta, F. Estrada, P. Wils, and S. Süsstrunk, "Salient region detection and segmentation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5008 LNCS, pp. 66–75, 2008, doi: 10.1007/978-3-540-79547-6_7.

[55] A.   BenSaïda,   "Shapiro-Wilk   and   Shapiro-Francia   normality   tests."
https://www.mathworks.com/matlabcentral/fileexchange/13964-shapiro-wilk-and-shapiro-francia-normality-tests?focused=3823443&tab=function (accessed Mar. 20, 2018).

[56] MathWorks,   "Statistics   and   Machine   Learning   Toolbox,"   2018.
https://www.mathworks.com/products/statistics.html (accessed Mar. 20, 2018).

[57] Mathworks, "Bioinformatics Toolbox." https://www.mathworks.com/products/bioinfo.html (accessed Mar. 20, 2018).

[58] Y. George, M. Aldeen, and R. Garnavi, "Skin Hair Removal for 2D Psoriasis Images," in *2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, Nov. 2015, pp. 1–8, doi: 10.1109/DICTA.2015.7371308.

[59] D. Stutz, A. Hermans, and B. Leibe, "Superpixels: An evaluation of the state-of-the-art," *Comput. Vis. Image Underst.*, vol. 166, pp. 1–27, 2018, doi: 10.1016/j.cviu.2017.03.007.

[60] G. Wang, "A perspective on deep imaging," *IEEE Access*, vol. 4, pp. 8914–8924, 2016, doi: 10.1109/ACCESS.2016.2624938.

[61] F. Henrique Schuindt da Silva, "Deep Learning for Corpus Callosum Segmentation in Brain Magnetic Resonance Images," 2018.

[62] T. Volkenandt, S. Freitag, and M. Rauscher, "Machine Learning Powered Image Segmentation," *Microsc. Microanal.*, vol. 24, no. S1, pp. 520–521, 2018, doi: 10.1017/s1431927618003094.

[63] A. Işın, C. Direkoğlu, and M. Şah, "Review of MRI-based Brain Tumor Image Segmentation Using Deep Learning Methods," *Procedia Comput. Sci.*, vol. 102, no. August, pp. 317–324, 2016, doi: 10.1016/j.procs.2016.09.407.

[64] R. Millioni, S. Sbrignadello, A. Tura, E. Iori, E. Murphy, and P. Tessari, "The inter- and intra-operator variability in manual spot segmentation and its effect on spot quantitation in two-dimensional electrophoresis analysis," *Electrophoresis*, vol. 31, no. 10, pp. 1739–1742, 2010, doi: 10.1002/elps.200900674.

[65] T. C. M. Lee and M. Fan, "Variants of seeded region growing," *IET Image Process.*, vol. 9, no. 6, pp. 478–485, 2014, doi: 10.1049/iet-ipr.2014.0490.

[66] J. Fan, R. Wang, S. Li, and C. Zhang, "Automated cervical cell image segmentation using level set based active contour model," *2012 12th Int. Conf. Control. Autom. Robot. Vision, ICARCV 2012*, vol. 2012, no. December, pp. 877–882, 2012, doi: 10.1109/ICARCV.2012.6485273.

[67] Y. J. Kim, S. H. Lee, C. M. Park, and K. G. Kim, "Evaluation of semi-automatic segmentation methods for persistent ground glass nodules on thin-section CT scans," *Healthc. Inform. Res.*, vol. 22, no. 4, pp. 305–315, 2016, doi: 10.4258/hir.2016.22.4.305.

[68] H. R. Roth *et al.*, "Deep learning and its application to medical image segmentation," pp. 1–6, 2018, doi: 10.11409/mit.36.63.

[69] X. Zhou *et al.*, "Performance evaluation of 2D and 3D deep learning approaches for automatic segmentation of multiple organs on CT images," in *Medical Imaging 2018: Computer-Aided Diagnosis*, Feb. 2018, vol. 10575, p. 83, doi: 10.1117/12.2295178.

[70] D. Shen, G. Wu, and H.-I. Suk, "Deep Learning in Medical Image Analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, no. 1, pp. 221–248, 2017, doi: 10.1146/annurev-bioeng-071516-044442.

[71] K. Suzuki, "Overview of deep learning in medical imaging," *Radiol. Phys. Technol.*, vol. 10, no. 3, pp. 257–273, 2017, doi: 10.1007/s12194-017-0406-5.

[72] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing*, vol. 187, pp. 27–48, 2016, doi: 10.1016/j.neucom.2015.09.116.

[73]   A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Adv. Neural Inf. Process. Syst.*, pp. 1–9, 2012, doi: http://dx.doi.org/10.1016/j.protcy.2014.09.007.

[74]   A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, P. Martinez-Gonzalez, and J. Garcia-Rodriguez, "A survey on deep learning techniques for image and video semantic segmentation," *Appl. Soft Comput. J.*, vol. 70, pp. 41–65, 2018, doi: 10.1016/j.asoc.2018.05.018.

[75]   O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9351, pp. 234–241, 2015, doi: 10.1007/978-3-319-24574-4_28.

[76]   F. Milletari, N. Navab, and S. A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," *Proc. - 2016 4th Int. Conf. 3D Vision, 3DV 2016*, pp. 565–571, 2016, doi: 10.1109/3DV.2016.79.

[77]   G. Csurka, D. Larlus, and F. Perronnin, "What is a good evaluation measure for semantic segmentation?," *BMVC 2013 - Electron. Proc. Br. Mach. Vis. Conf. 2013*, 2013, doi: 10.5244/C.27.32.

[78]   Y. Xu, Y. Wang, J. Yuan, Q. Cheng, X. Wang, and P. L. Carson, "Medical breast ultrasound image segmentation by machine learning," *Ultrasonics*, vol. 91, no. July 2018, pp. 1–9, 2019, doi: 10.1016/j.ultras.2018.07.006.

[79]   H. Chen, Q. Dou, L. Yu, J. Qin, and P.-A. Heng, "VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images," *Neuroimage*, vol. 170, no. April 2017, pp. 446–455, Apr. 2018, doi: 10.1016/j.neuroimage.2017.04.041.

[80]   A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool," *BMC Med. Imaging*, vol. 15, no. 1, 2015, doi: 10.1186/s12880-015-0068-x.

[81]   H. Costa, G. M. Foody, and D. S. Boyd, "Supervised methods of image segmentation accuracy assessment in land cover mapping," *Remote Sens. Environ.*, vol. 205, no. December 2016, pp. 338–351, 2018, doi: 10.1016/j.rse.2017.11.024.

[82]   M.-P. Dubuisson and A. K. Jain, "A modified Hausdorff distance for object matching," no. 1, pp. 566–568, 2002, doi: 10.1109/icpr.1994.576361.

[83]   N. I. of H.-C. Center, "Chest X-Ray NIHCC," 2017. https://nihcc.app.box.com/v/ChestXray-NIHCC (accessed Nov. 10, 2019).

[84]   T. M. I. of T. (MIT)'s L. for C. Physiology, "MIMIC-Chest X-Ray Database (MIMIC-CXR)." https://physionet.org/content/mimic-cxr/2.0.0/ (accessed Nov. 30, 2019).

[85]   K. Yan, X. Wang, L. Lu, and R. M. Summers, "DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning," *J. Med. Imaging*, vol. 5, no. 03, p. 1, Jul. 2018, doi: 10.1117/1.JMI.5.3.036501.

[86]   A. F. Fotenos, A. Z. Snyder, L. E. Girton, J. C. Morris, and R. L. Buckner, "Normative estimates of cross-sectional and longitudinal brain volume decline in aging and AD," *Neurology*, vol. 64, no. 6, pp. 1032–1039, Mar. 2005, doi: 10.1212/01.WNL.0000154530.72969.11.

[87]   M. S. Badea, I. I. Felea, L. M. Florea, and C. Vertan, "The use of deep learning in image segmentation, classification and detection," no. Section II, pp. 1–5, 2016, [Online]. Available: http://arxiv.org/abs/1605.09612.

[88]   N. Dhungel, G. Carneiro, and A. P. Bradley, "Deep Learning and Structured Prediction for the Segmentation of Mass in Mammograms," in *Navab N., Hornegger J., Wells W., Frangi A. (eds) Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science*, vol. 9349, 2015, pp. 605–612.

[89]   P. Moeskops *et al.*, "Deep learning for multi-task medical image segmentation in multiple modalities," *Lect.*

Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9901 LNCS, no. October, pp. 478–486, 2016, doi: 10.1007/978-3-319-46723-8_55.

[90] Q. Dou *et al.*, "3D deeply supervised network for automated segmentation of volumetric medical images," *Med. Image Anal.*, vol. 41, pp. 40–54, 2017, doi: 10.1016/j.media.2017.05.001.

[91] G. Wang *et al.*, "Interactive Medical Image Segmentation Using Deep Learning with Image-Specific Fine Tuning," *IEEE Trans. Med. Imaging*, vol. 37, no. 7, pp. 1562–1573, 2018, doi: 10.1109/TMI.2018.2791721.

[92] M. Havaei *et al.*, "Brain tumor segmentation with Deep Neural Networks," *Med. Image Anal.*, vol. 35, pp. 18–31, 2017, doi: 10.1016/j.media.2016.05.004.

[93] T. A. Ngo, Z. Lu, and G. Carneiro, "Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance," *Med. Image Anal.*, vol. 35, pp. 159–171, 2017, doi: 10.1016/j.media.2016.05.009.

[94] F. Milletari *et al.*, "Hough-CNN: Deep learning for segmentation of deep brain regions in MRI and ultrasound," *Comput. Vis. Image Underst.*, vol. 164, pp. 92–102, 2017, doi: 10.1016/j.cviu.2017.04.002.

[95] J. L. Prince *et al.*, "Parallel deep neural networks for endoscopic OCT image segmentation," *Biomed. Opt. Express*, vol. 10, no. 3, p. 1126, 2019, doi: 10.1364/boe.10.001126.

[96] Z. Jia, X. Huang, E. I. C. Chang, and Y. Xu, "Constrained Deep Weak Supervision for Histopathology Image Segmentation," *IEEE Trans. Med. Imaging*, vol. 36, no. 11, pp. 2376–2388, 2017, doi: 10.1109/TMI.2017.2724070.

[97] Z. Zhao, L. Yang, H. Zheng, I. H. Guldner, S. Zhang, and D. Z. Chen, "Deep Learning Based Instance Segmentation in 3D Biomedical Images Using Weak Annotation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11073 LNCS, pp. 352–360, 2018, doi: 10.1007/978-3-030-00937-3_41.

[98] M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," in *European Conference on Computer Vision (ECCV) 2014*, 2014, pp. 818–833, doi: https://doi.org/10.1007/978-3-319-10590-1_53.

[99] K. Simonyan, "Deep Inside Convolutional Networks : Visualising Image Classification Models and Saliency Maps arXiv : 1312 . 6034v2 [cs . CV] 19 Apr 2014," pp. 1–8, 2013.